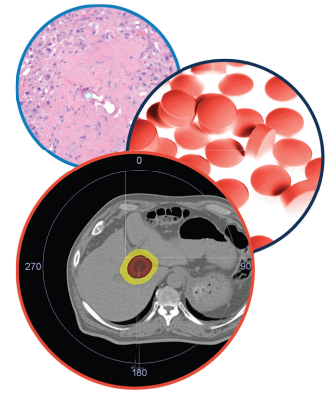
Perspective



Hepatic Oncology

Risk assessment criteria in liver transplantation for hepatocellular carcinoma: proposal to improve transplant oncology

John C McVey[1,4], Kazunari Sasaki[2] & Daniel J Firl*[,3]

[1]Cleveland Clinic Lerner College of Medicine of Case Western Reserve University, Cleveland, OH 44113, USA
[2]Department of General Surgery, Digestive Disease & Surgery Institute, Cleveland Clinic, Cleveland, OH 44113, USA
[3]Department of Surgery, Duke University School of Medicine, Durham, NC 27705, USA
[4]Gastrointestinal & Thoracic Malignancy Section, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA
*Author for correspondence: daniel.firl@duke.edu

Liver transplantation for hepatocellular carcinoma has proved to be a highly effective cure if the right patient can be selected. Milan criteria has traditionally guided physicians toward appropriate liver allocation but changes in clinical practice, patient populations and recent developments in biomarkers are decreasing Milan criteria's utility. At the same time, the literature has flooded with a diversity of new criteria that demonstrate strong predictive power and are better suited for current clinical practice. In this article, the utility of newly proposed criteria will be reviewed and important issues to improve future criteria will be addressed in hopes of opening a discussion on how key questions surrounding criteria for liver transplantation of hepatocellular carcinoma can be answered.

First draft submitted: 7 February 2020; Accepted for publication: 23 June 2020; Published online: 24 July 2020

Worldwide, liver cancer accounts for the fourth most frequent cause of cancer-related death and the sixth-highest incidence among all cancer types [1]. Hepatocellular carcinoma (HCC) makes up the majority of all primary liver cancers and has been largely considered a surgically treated malignancy because of poor availability of effective systemic therapy. To date, liver transplantation (LT) remains the best chance at cure because the carcinogenic background in which the malignancy formed is removed [2]. Even though LT represents the gold standard therapy, it is only available to a subset of patients with few and small tumors. In general, these patients are determined by a cutoff called the Milan criteria (MC) which for the first time demonstrated that patients with one tumor up to 5 cm or three tumors totaling less than 3 cm had acceptable survival and recurrence rate post-LT [3]. Since its inception in 1996, MC has become the cornerstone of most liver allocation systems worldwide and ushered in an era of assessing oncologic transplant candidacy through preset criteria. While MC was the first and is still the most common criteria for determining LT candidacy, its utility has come under scrutiny. At the same time, an immense number of newly proposed criteria have flooded the literature. In this article, the authors will summarize the current utility of MC, investigate the utility of other important criteria, provide a summary of the data for a selected group of proposed criteria and finally review considerations and key questions to improve the utility of future criteria. While there are a number of well written articles that assess a wide verity of topics related to HCC risk assessment in LT such as their use around the world [4] and optimal metrics for LT of HCC [5]. The goal of this article is to leave the reader with a better understanding of the current utility of proposed criteria for HCC and enrich a discussion on how the field can address critical questions surrounding these tools, to gain wider acceptance.

Milan criteria & its diminished utility

Prior to 1996, there was much debate about the utility of LT for primary liver cancers with early studies showing poor overall survival and high recurrence rates [6–8]. In these studies, however, there was a subset of patients that responded very well to LT and were effectively cured of their disease for an extended period of time. This

Future Medicine

finding indicated that maybe LT for HCC was a viable treatment option if the right patient could be selected. One of the biggest challenges facing physicians attempting to transplant any oncology patient is the decreased immunosurveillance from the use of required immunosuppression post-LT leading to high rates of recurrence. This made selecting the right patient even more difficult in an era where little information was available on tumor markers and biology. It was not until an Italian group from the National Cancer Institute at Milan conducted a prospective trial on 48 cirrhotic patients with unresectable tumors that the concept of using tumor morphology (tumor size and number) was definitively proven to be an effective selection criteria for LT of HCC [3]. In this study, Mazzaferro *et al.* provided explant pathology which showed that patients with one tumor up to 5 cm or three tumors, each totaling no more than 3 cm had superior recurrence rates and overall survival compared with patients that had larger and/or more numerous tumors. This study established a simple criteria which could be applied worldwide and to this day is still widely accepted as the benchmark for determining LT candidacy for patients with HCC.

The significance of MC is highlighted by the fact that the original study by Mazzaferro *et al.* has been cited well over 6000-times and that MC has been adopted by roughly 95% of countries as the main criteria used to determine LT candidacy [4]. Even with all of its success, MC has its short comings. Only 5 years later, a single center retrospective analysis showed that patients beyond MC could be transplanted with acceptable outcomes suggesting that MC is too restrictive [9]. Clinical practice has also changed over the past 25 years. Most notably, there has been an ever-increasing penetration of bridging locoregional therapy (LRT). Following LRT, some large tumors lie quiescent whereas some smaller ones persist and progress despite similar treatments. These disparities have led to the natural understanding that tumor morphology alone is not predictive of tumor biology [10,11]. There has also been a shift in etiology leading to cirrhosis and eventual HCC. In the western world, the primary etiology driving cirrhosis was HCV, but with increases in preventative measures and improvements in anti-viral therapy, HCV rates have declined. Rises in obesity and metabolic syndrome including nonalcoholic fatty liver disease are now projected to increase and may surpass other background etiologies leading to HCC [1]. In parallel, there has been a greater understanding of biomarkers such as α-fetoprotein (AFP) and immune cells such as systemic neutrophil activation/proliferation or tumor infiltrating T and B cells and their predictive power in HCC [12–17]. Finally, in many countries that use MC, there are disparities in the allocation of precious livers between HCC and non-HCC patients with an unfair advantage given to patients that are within MC [18,19]. This is especially true in the USA where the United Network of Organ Sharing (UNOS) has continuously attempted to update HCC LT policy but has so far failed to abolish transplant discrepancies [20]. The most recent of these rules was the addition of an AFP cutoff which excludes patients from exception points if their serum AFP is greater than 1000 unless there is a response to LRT. These patchworks of policies, however, have made a system that is more cumbersome and entirely reactionary rather than proactively thought out.

At this point, it is widely accepted that MC is not the best selection criteria and should be replaced [21,22]. For the reasons listed above, there has been great interest and number of new criteria in the literature attempting to safely extend beyond MC. The newly proposed criteria vary widely in the risk parameters they use, outcomes they assessed and information they provide giving the field a diversity of novel risk assessment tools to choose from. Outside of a few select countries, however, these criteria have not gained wide acceptance even though they are more current and show superior predictive power than MC. Furthermore, many of the proposed allocation criteria are not generalizable or have not undergone external validation which limits their utility and thus their application.

Proposed criteria & their utility

Over the past 25 years, our field has made huge advances in the discovery of predictive markers other than tumor morphology which has improved the utility of newly proposed criteria. As shown in Figure 1, other factors such as AFP levels [13,23–33], immune cell populations [26,34], biopsy results [23,30], Des-gamma-carboxy prothrombin [35] and model for end-stage liver disease (MELD) [28,31,32] have all been applied to improve prognosticating how an HCC patient will fair before or after LT. The most well-known and commonly used metric is AFP (Figure 1). AFP levels have been known to be elevated in patients with HCC for many years [36] and clearly convey a survival benefit in patients with lower concentrations [37]. This biomarker, however, is not predictive for all patients, especially those with small tumors and well-compensated cirrhosis [38]. Further, even some tumors with bad biology do not produce AFP which is deleterious to many scoring systems. Other biomarkers such as various immune cell populations, which can easily be assessed through common labs, have also been tested and demonstrate predictive value [39]. It is important to note, however, that there is question if systemic inflammatory markers are really predictive of tumor

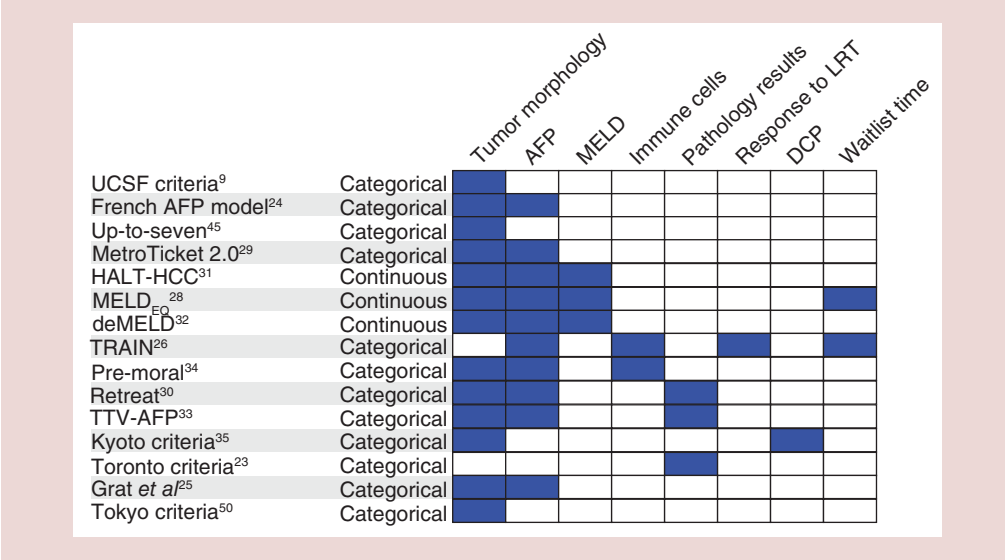


Figure 1. Summary of proposed criteria: characteristics of proposed criteria including type of output and metrics used.

deMELD: Dropout equivalent model for end-stage liver disease; HALTHCC: Hazard associated with liver transplantation for hepatocellular carcinoma; LRT: Locoregional therapy; MELD: Model for end-stage liver disease; MELD$_{EQ}$: Equivalent model for end-stage liver disease; Moral: Model of recurrence after liver transplantation; Retreat: Risk estimation of tumor recurrence after transplant; TRAIN: Time, radiological response, alpha-fetoprotein and inflammation score; TTV-AFP: Total tumor volume and alpha-fetoprotein; UCSF: University of California San Francisco.

biology and not confounded by other factors affecting systemic inflammation [40–42]. More useful populations of tumor-infiltrating lymphocytes or activation profiling are unfortunately more center dependent and more difficult to reproduce or generalize, which limits the applicability of immune cells as a whole in improving scoring system utility. MELD score itself has also been included in a number of risk assessment tools. Including MELD may provide important information on the overall fitness of the patient since many HCC patients develop their disease in the background of cirrhosis. That being said, the majority of HCC patients are well compensated and transplanted with low MELD scores, potentially diminishing the significance of this marker in risk assessment. Finally, some risk tools use markers such as pathology results and Des-gamma-carboxy prothrombin, which are known to correlate with LT outcomes, but may not be part of standard of care at many institutions and will thus limit their generalizability [23,43].

Response to LRT is becoming an increasingly recognized important marker of patient's risk of recurrence and survival post-LT [21]. With the organ shortage, the majority of patients undergoing LT in competitive regions end up receiving some form of neo-adjuvant therapy prior to LT. The observed response to LRT (in tumor biomarkers, such as AFP, or morphology, such as the modified Response Evaluation Criteria in Solid Tumors [mRECIST]) may provide valuable information on tumor behavior. One study has shown that patients originally outside of MC that are down staged with LRT to MC in, had similar post-LT outcomes to those that were never outside of MC [44]. The inclusion of response to LRT is being used in different ways for different risk assessment criteria. For example, TRAIN score includes radiologic response to LRT as an independent variable in its risk calculation [26]. More recently, a study showed that the inclusion of mRECIST, post neo-adjuvant therapy, in MetroTicket2.0 improved its predictive value [45]. Other scores, such as Hazard Associated with Liver Transplantation for HCC (HALT-HCC), do not included response to LRT as an outright variable in its calculation of risk, but instead found other ways to integrate the information into assessing post-LT risk [46]. In all of these studies, patients that respond well to LRT generally have better outcomes post-LT, highlighting its significance in predicting risk. It is important to note, predictive tools in other forms of cancer have also used response post-treatment and demonstrated improved predictive value over single time point analysis [47].

With many tools to choose from, the question of what exactly to include in a criteria to best determine LT candidacy is up for debate. It is widely accepted that at minimum there must be a combination of tumor morphology and tumor biology (most commonly assessed through serum markers such as AFP) [5,48]. This claim is supported by the fact that the vast majority of newer proposed scores include a combination of tumor morphology and biology

while older scores (Up-to-Seven, University of California San Francisco [UCSF] and Tokyo Criteria) only include tumor morphology (Figure 1). To date, however, the authors do not know of a recent study that has, as its primary purpose, compared the predictive value of various newly proposed criteria with different combinations of tumor metrics against each other. Comparative utility analyses are sometimes provided in the original study describing a new tool, but this method is biased since the dataset used to compare the tool is often the same one used to train or recalibrate the proposed score. A study using an untested dataset with the primary end point of determining the most predictive proposed tool is lacking and seriously needed to effectively determine which proposed score contains the right combination of patient/tumor metrics.

The other major factor affecting the utility of newly proposed criteria is the information the score provides the physician. The output of various criteria is either a categorical [9,23–26,29,30,34,35,49–51] (e.g., transplant or do not transplant) or else a continuous output [13,28,31,32] (similar to MELD-Na) (Figure 1). There are advantages and disadvantages to both approaches; for example, having a categorical output makes the decision to transplant much easier because a yes or no answer is provided. However, the granularity of how one patient's risk compared with another is lost. Having a continuous output maintains this granularity but may make it more difficult to integrate into current allocation systems that often provide MELD-Na exception points based off of a transplant/do not transplant categorization. To highlight these issues, a few studies have investigated how newly proposed tools could fit into current liver transplant policy while still providing as much information to providers as possible.

Two proposed tools that provide a continuous output (deMELD and $MELD_{EQ}$) have investigated how a continuous HCC score could integrate into current liver allocation systems using MELD exception points [28,32]. In both of these retrospective studies, tumor morphology, AFP and MELD were used to predict risk of all cause waitlist dropout for HCC patients (Figure 1). The predicted risk was then matched with the predicted risk of 3-month dropout calculated by MELD in non-HCC patients to provide a new MELD score. Both studies showed excellent predictive value for waitlist dropout in HCC patients (deMELD C-statistic = 0.73 and $MELD_{EQ}$ C-statistic = 0.74). They were also able to demonstrate a potential decrease in the discrepancy between HCC and non-HCC liver allocation. While the focus of these studies was on pre-LT outcomes, post-LT survival was also assessed and increases in both scores were associated with risk of poorer post-LT outcomes [52]. Its critical to remember however, that this would not be possible with any tool that yields a categorical output. Thus, these studies highlight the potential of matching HCC and non-HCC risk of dropout to create a new continuous 'HCC MELD' score that can integrate into current liver allocation. There are unsolved issues with this methodology. For example, it is not hard to conceive a situation where a patient has a single small tumor and low AFP value with very high MELD score. In this case, the patient's true MELD score would be higher than their 'HCC MELD' score and thus put them at a disadvantage. The question of whether to transplant the patient based on their MELD versus 'HCC MELD' is yet to be determined. Furthermore, the LT community has been transplanting patients based off of a transplant/do not transplant scheme for many years and there is thus a lot of inertia that must be overcome if a continuous tool is ever going to be taken seriously. Our suggestion is to hold a conference with all important stakeholders and field experts where a consensus can be reached on needed future studies and terminology to advance the field in coming years.

Summary of the data for a select group of proposed criteria

One of the first proposed criteria after MC was the University of California San Francisco (UCSF) criteria. UCSF criteria has gained popularity around the world and has even replaced MC in some countries as the main tool determining liver allocation for HCC [4]. The original UCSF study investigated 70 patients transplanted at their single center between 1988–2000 and found that patients with a single lesion less then 6.5 cm or up to three lesions less than 4.5 cm each with a total tumor diameter less then 8 cm had acceptable outcomes post-LT. In this study, tumor size and number were determined by pathological analysis and not pre-LT imaging. This study found that an AFP level >1000 ng/ml was predictive of post-LT mortality, but the authors did not include it in their tool. For patients that met UCSF criteria, the 5-year survival rate post-LT was 80% compared with <50% in the group that exceeded UCSF criteria [9]. This finding was then validated by a number of other external studies that clearly demonstrated its superior utility over MC [53–55]. The use of this expanded criteria has also been estimated to increase the indication of LT for HCC patients by 5–20% [48]. While this study has gained popularity, it still suffers many of the same issues that face MC. For example, similar to MC, it was created in an era that had dramatically different clinical practice, utilized nonpre-LT-accessible pathological results for tumor size and number and does

Table 1. Each risk metric reviewed in this article and the calculation used to determine risk.

Risk metric	Risk calculation	Risk cutoff
UCSF criteria	One lesion <6.5 cm or 3 lesions <4.5 cm each with a total tumor diameter <8 cm	Meets criteria
French AFP model	Largest tumor diameter (<3 cm = 0, 3–6 cm = 1 and >6 = 4 points) + number of lesions (1–3 = 0 and >4 = 2 points) + AFP level (<100 = 0, 100–1000 = 2 and >1000 = 3 points)	Sum of points is less than or equal to two
Up-to-Seven	Number of lesions + largest lesion size (cm)	Sum of calculation is less than seven
MetroTicket 2.0	Number of lesions + largest lesion size (cm) <7 and AFP <200 or number of lesions + largest lesion size (cm) <5 and AFP <400 or number of lesions + largest lesion size (cm) <4 and AFP <1000	Meets criteria
HALT-HCC	(1·27 × Tumor Burden Score) + (1·85 × lnAFP) + (0·26 × MELD-Na)	Continous score without cutoff
$MELD_{EQ}$	(1:143 × Lab MELD) + (1.324 × log AFP) + (1.438 × lesion number) + (1.194 × lesion size [cm]) + c(t) (where c(t) = -2/0.146 if WL time <6 months or -1/0.146 if WL time >6 months)	Continous score without cutoff
deMELD	37.8 + (1.9 x MELD) + (5.9 if lesion number >1) + (5.9 if AFP >400) + (21.2 if lesion size >1 cm)	Continous score without cutoff
TRAIN	0.988 (if mRECIST is PD) + 0.838 (if AFP slope >15 ng/ml/month) + 0.452 (if NLR >5) - (0.03 x WL time [months])	Sum of the points is less than one
Pre-Moral	3 (if largest lesion >3 [cm]) + 6 (if NLR >5) + 4 (if AFP >200)	0–2 is low risk, 3–6 is medium risk, 7–10 is high risk and 11–13 is very high risk
Retreat	AFP at LT (0–20 = 0, 21–99 = 1, 100–999 = 2 and >1000 = 3 points) + microvascular invasion (2 points if present) + sum of largest lesion and number of lesions (0 = 0, 1.1–4.9 = 1, 5–9.9 = 2 and >10 = 3 points)	No cutoff, risk is stratified by number of points obtained
TTV-AFP	Total lesion volume <115 cm^3 and AFP <400	Meets criteria
Kyoto criteria	Total lesion number <11 and all lesions <5 cm and DCP <400 mAU/ml	Meets criteria
Toronto criteria	No systemic symptoms and lesions not poorly differentiated	Meets criteria
Grat *et al.*	Meets UCSF and Up-to-Seven and has an AFP <100	Meets criteria
Tokyo criteria	Five lesions with no lesion >5 cm	Meets criteria

Also included is the cutoff point used in the original study.
DCP: Des-gamma-carboxy prothrombin; HALT-HCC: Hazard associated with liver transplantation for hepatocellular carcinoma; LT: Liver transplantation; MELD: Model for end-stage liver disease; mRECIST: Modified response evaluation criteria in solid tumors; NLR: Neutrophil to lymphocyte ratio; PD: Progressive disease; TRAIN: Time, radiological response, alpha-fetoprotein and inflammation score; TTV-AFP: Total tumor volume and alpha-fetoprotein; UCSF: University of California San Francisco; WL: Wait list.

not account for important markers of tumor biology. Despite these caveats, a later validation study using a more modern cohort and pre-LT imaging was still able to establish UCSF criteria's utility over MC [56].

In 2009, the Milan group proposed an update to MC called the Up-to-Seven criteria (sometimes referred to as MetroTicket). In this study, data entered into an online survey from centers around the world was utilized to investigate outcomes of patients transplanted beyond MC. At the end of the recruitment period, the study consisted of 36 transplant centers with a total of 1556 patients transplanted from 1984–2006. It was found that the 5-year overall survival rate for patients that met the Up-to-Seven criteria (sum of number of lesions plus the size of the largest tumor is less then 7) was 71.2% which was similar to MC and other criteria at the time [49]. This can be compared with the group of patients outside of Up-to-Seven which had a 5-year overall survival of only 48.1%. Further validation of this criteria showed that it was generalizable and predictive of recurrence post-LT [57,58]. In 2018, Up-to-Seven (MetroTicket) was updated to MetroTicket 2.0. This updated version was created using a competing risk regression model analyzing 1018 patients from three centers Italy and then validated with 341 patients from the same centers. The study found that the sum of tumor number and largest tumor size with the addition of AFP was very predictive of overall survival, HCC specific survival and recurrence free survival [29]. It is important to note that MetroTicket 2.0 was first a continuous score but was adjusted to provide a dichotomous output later in the study (Up-to-Seven and AFP <200, Up-to-Five and AFP <400 or Up-to-Four and AFP <1000). The study also showed that MetroTicket 2.0 was superior at predicting HCC specific survival (C-statistic = 0.72) compared with AFP model, UCSF criteria, Up-to-Seven, MC and Shangai-Fudan.

The AFP model (sometimes referred to as the French AFP model) was one of the first criteria to integrate AFP with tumor morphology. The AFP model was created using tumor size, number and AFP and a training dataset consisting of 597 HCC patients who underwent LT between 1988 and 2001 at 16 centers across France. The model was then validated on another dataset consisting of 474 HCC patients transplanted between 2003 and 2004 at 21 centers in France. The AFP model consists of a point system where patient with a sum of 2 points or less are considered low risk and greater than 2 points are high risk (Table 1). The study found that patients considered low risk by the AFP model had a 5-year overall survival of 67.8% (vs 47.5% p = 0.002) and 5-year recurrence rate of 8.8% (vs 50.6% p < 0.001) [24]. The results of this study have been validated in other patient populations thus

proving its generalizability [59]. Shortly after the original study was published, France adopted the AFP model as its HCC criteria for determining liver allocation. This adoption of a combined morphology/biology tool represents a model for nations interested in rationally advancing patient care and is worthy of consideration and implementation for other tools worldwide.

While the other criteria listed above have relied on tumor morphology and AFP to produce a categorical output, the HALT-HCC is one of a few criteria that uses tumor morphology, AFP and MELD to produce a continuous score assessing risk. HALT-HCC was originally estimated from a retrospective cohort of 420 patients transplanted between 2002 and 2014 at a large single center in the USA and then validated using 13,717 patients from the Scientific Registry of Transplant Recipients. This tool showed excellent stratification of post-LT overall survival and superior predictive power over all other proposed scores (despite a relatively lower C-statistic = 0.68 which may be related to a highly heterogeneous study population). Further analysis showed that close to 30% of patients transplanted outside of MC had similarly good prognosis to MC-in patients and surprisingly, 21% of patients MC-in had similarly poor prognosis to MC-out patients based off of HALT-HCC scores [31]. In a follow-up study, HALT-HCC was recalibrated with 4089 patients from 16 centers around the world and was found to demonstrate superior predictive power for post-LT HCC recurrence as well as predictive explant pathological features like microvascular invasion and poorly differentiation tumor component [60]. In a third study, response to LRT based off of longitudinal HALT-HCC measure was tested to determine if it could predict waitlist dropout, post-LT recurrence and overall survival. Results showed that HALT-HCC outperformed the gold standard response criteria, mRECIST, in all three of these outcomes thus demonstrating its potential utility as an important prognostic tool for determining liver allocation priority for patients on the waiting list [46].

Considerations to improve the utility of future scores

Even though there are many proposed criteria available, very few have gained acceptance and even fewer are being utilized by governing organizations. Two main transplant governing bodies, UNOS (USA) and Eurotransplant (seven countries in Europe), still endorse MC as the main criteria for determining whether an HCC patient can receive a LT. There have, however, been a few countries that have adopted other criteria [4]. As stated above, France created and has used the AFP model in conjunction with tumor, node and metastasis (TNM) classification in the American Joint Committee on Cancer staging (AJCC) to determine MELD exception points [61]. Other countries such as Russia, Australia/New Zealand and India have adopted the UCSF criteria. Why more countries have not accepted new criteria is likely multifactorial. There are, however, two major issues that should be addressed and will likely greatly increase the utility of future criteria. First, there needs to be a discussion on who receives allocation priority within HCC patients and two, there needs to be improved validation and simulation tools to assess and confirm the validity of a proposed criteria to address the complex challenges facing LT for HCC.

Mazzaferro *et al.* elegantly stated that one of the major obstacles facing a smooth HCC liver allocation system is the availability of appropriate instruments to accurately determine how 'oncologically sick' a patient is or how a tumor will respond to a given therapy [21]. As discussed in this review article, newly proposed criteria are becoming increasing more powerful and providing a level of detailed information to a physician that was not possible even 10 years ago. This is especially true for continuous scores that can now allow for direct comparisons between patients. The problem facing the LT community in 1996 can thus be seen from a different perspective. Instead of determining who should not be given a liver based off of poor outcomes, it may not be long before we can ask "who should not receive a transplant because they are not sick enough?". With advances in LRT and large discrepancies between allocation to HCC and non-HCC patients, this is becoming an even more important question that the LT community must answer. Having an agreed upon strategy for liver allocation (sickest first vs greatest utility vs hybrid) will guide future criteria on what metrics to include and the amount of weight to place on each. There are clearly disadvantages and advantages to any allocation strategy, but a community wide discussion is really needed to best determine how future criteria are utilized.

Virtually all proposed criteria discussed in this review article have relied on retrospective analysis to determine their prognostic power and potential utility. While this methodology is easy and has provided some insight into the utility of new criteria, it falls short of addressing the true complexity of the modern HCC allocation system. Using this form of study, it is impossible for a researcher to ask if a new criteria could decrease the discrepancy between HCC and non-HCC patients or how a criteria will change allocation across multiple regions. As we have learned from MC, these are important questions that need to be answered. In non-HCC allocation, there are a number of tools that have helped researchers answer these nuanced questions and effectively implement new policy. For

example, in the USA, UNOS developed a computer modeling system called liver simulated allocation model using Scientific Registry of Transplant Recipients data that allows researchers to change transplant allocation policy and see how organ offers, acceptance and waitlist times are affected. This program has been used to actually change allocation policy in the United States [62,63]. Other independent researchers have also created other computer modeling tools that have shaped UNOS policy [64]. Interestingly, none of these tools include information on HCC patients. We suggest modification of the UNOS tool liver simulated allocation model to allow more complex cancer allocation changes to be modeled. This simulation environment would then provide a universal framework through which all proposed tools can be compared prior to the significant investment of time and resources represented by allocation policy change. Furthermore, almost no prospective clinical trials are run testing how newly proposed criteria will affect allocation. For a new criteria to be truly tested, much work needs to be done in the area of how we actually test and confirm its utility. Based off of the adoption rate of newly proposed criteria, it is clear that retrospective studies are no longer sufficient to convince the transplant community of their utility. Only when we can actually simulate liver allocation will the true effects of a proposed criteria be realized.

Another challenge all future scores will have to address is the differences in liver transplantation between countries. For example, in Asian countries there is a dominance of living donor liver transplantation over deceased donor, which is more common in western countries [4]. There are also key differences in background etiology leading to HCC by region [2]. A metric designed for one region of the world will likely underperform in a region different from the one it was trained on. Using computer modeling or reevaluating newly proposed metrics on cohorts representative of the region will be crucial for determining how a metric will perform. Metrics can then be adjusted to that region to account for differences. It is important to note that having a continuous score makes this much easier because 'mathematical weights' can be adjusted for each variable within the metric to reflect their importance by region. This also highlights another value of computer modeling since adjustments can be done before a metric is tested in a prospective trial.

Finally, central to an understanding of how future criteria can be improved is a deep respect for the questions these tools are trying to answer. LT as a whole has long been a field intertwined with the ethics that accompany the decision of whose life is worth saving with a limited resource. LT for HCC is no different. All future criteria will have to address this question in a manner that is fair, evidence based and accounts for the opinions of all the important stakeholders. As many studies have previously shown, there is a group of patients that will not benefit from LT [65]. This group is generally high risk and are better suited to receive other forms of therapies such as LRT or systemic therapies that can provide prolongation of life without the use of a precious resource [66,67]. As stated above, it is likely that future LT criteria will also be able to identify patients that are low risk for tumor progression. In this case, these patients might be better suited for neo-adjuvant therapies and do not need expeditious transplantation. Also central to the ethics of future HCC LT criteria is its integration into the MELD system for non-HCC patients. As we have seen with exception point systems, there is a potential for one group to be at a disproportionate advantage over the other [19,20]. Building a criteria that can be modified with adjustable mathematical weights will likely abolish this discrepancy and provide fair allocation for all. Until organs are no longer a limited resource, the ethical questions asked by LT for HCC will have to be addressed with agreed upon powerful risk assessment criteria.

Conclusion

Advances in the locoregional treatment of HCC have demonstrated that therapies can dramatically improve patient's lives without having to undergo LT [66,68,69]. While these advances are impressive, they are still far from accomplishing what LT can provide the appropriate HCC patient. For the time being, this has cemented LT as the best possible cure for HCC and it is thus imperative that as many HCC patients as possible have access to a transplant. Crucial to this accessibility, however, is a deep understanding of who can actually benefit from such a procedure and who is better suited with less invasive options. MC was the first to take on this difficult task in an era where little resources were available to interrogate a tumor. It is commendable that MC has lasted as long as it has and integrated into so many liver allocation systems around the world. Much has changed, however, since MC was described and adopted. Changes in clinical practice, increased utilization of biomarkers and an understanding that MC withholds many patients from a cure have all made this outdated criteria obsolete. As highlighted by this review, the LT community knows this and has worked hard to produce and investigate an immense number of new criteria. The literature has shown that many of these new criteria are superior to MC, better equipped for current clinical practice and allow more patients access to a lifesaving procedure. Yet MC still remains dominant. With further clarification of HCC allocation priority and improved validation models, the utility of proposed tools will

likely be strengthened. These future considerations are important because the decision to replace MC is really up to us. We are privileged that many countries have granted the transplant community the unique right to self-govern. Our community has this right but with it a responsibility to take action in this, a new decade, to band together and hold consensus meetings, improve simulation tools and take serious action on a dramatic overhaul of the allocation system for LT for HCC. Until then, patients will continue to receive, or more importantly to not receive, life-saving liver transplantation based off of a system that we all know desperately needs improvement.

Future perspective

Liver transplantation for HCC will likely be much different 10 years from now then it is today. The wide acceptance that MC is no longer a sufficient risk assessment criteria will push researchers to develop powerful risk tools which integrate a diversity of markers that ultimately improve transplant selection. It is possible that a new continuous system, that steps away from exception point policy, will integrate seamlessly into the MELD system to minimize the discrepancies between HCC and non-HCC allocation. It is also possible that these new powerful risk assessment tools will allow for the creation of thresholds at the upper and lower bounds of risk. Candidates for LRT or novel neoadjuvant therapies will then be identified and precious organs will be conserved for those that need them most. We hope that this review highlights some of the steps that can be taken in the immediate time frame to improve the likelihood that these changes happen in the future.

Executive summary

- Milan Criteria was the first and still the most widely used criteria for determining hepatocellular carcinoma (HCC) liver transplant candidacy.
- Recently proposed criteria consider a wide variety of HCC factors outside of tumor morphology including serum biomarkers (α-fetoprotein and Des-gamma-carboxy prothrombin), immune cell correlates and response to locoregional therapies.
- The diversity of proposed metrics has demonstrated a number of options to replace or improve the utility of Milan Criteria.
- There is a need for new large and international studies which focus on establishing the most useful patient/tumor factors in untested datasets to minimize bias in determining the superior risk assessment tool.
- Computer modeling systems, such as liver simulated allocation model, should be updated to include HCC patients and HCC allocation factors (e.g., tumor size and number and α-fetoprotein level) so future risk assessment tools and policies can be tested and modeled.
- An added, often underappreciated benefit of continuous risk scoring tools is the potential to tell us who is not sick enough (or put another way, oncologically advanced enough) to receive a liver transplant for their low-risk cancer at this time.
- A consensus meeting with all important stakeholders is needed to determine future studies and terminology to advance the field in coming years.

Author contributions

JC McVey and DJ Firl wrote and prepared the manuscript. K Sasaki provided critical review.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

Reference

1. Villanueva A. Hepatocellular carcinoma. *N. Engl. J. Med.* 380(15), 1450–1462 (2019).
2. Forner A, Reig M, Bruix J. Hepatocellular carcinoma. *Lancet* 391(10127), 1301–1314 (2018).

3. 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.* 334(11), 693–699 (1996).
4. Tschuor C, Ferrarese A, Kuemmerli C, Dutkowski P, Burra P, Clavien PA. Allocation of liver grafts worldwide – is there a best system? *J. Hepatol.* 71(4), 707–718 (2019).
5. Mehta N, Yao FY. What are the optimal liver transplantation criteria for hepatocellular carcinoma? *Clin. Liver. Dis. (Hoboken)* 13(1), 20–25 (2019).
6. Haug CE, Jenkins RL, Rohrer RJ *et al.* Liver transplantation for primary hepatic cancer. *Transplantation* 53(2), 376–382 (1992).
7. Iwatsuki S, Gordon RD, Shaw BW Jr, Starzl TE. Role of liver transplantation in cancer therapy. *Ann. Surg.* 202(4), 401–407 (1985).
8. Olthoff KM, Millis JM, Rosove MH, Goldstein LI, Ramming KP, Busuttil RW. Is liver transplantation justified for the treatment of hepatic malignancies? *Arch. Surg.* 125(10), 1261–1266 (1990).
9. 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* 33(6), 1394–1403 (2001).
10. Morris PD, Laurence JM, Yeo D *et al.* Can response to locoregional therapy help predict longterm survival after liver transplantation for hepatocellular carcinoma? A systematic review. *Liver Transpl.* 23(3), 375–385 (2017).
11. Sheth RA, Patel MS, Koottappillil B *et al.* Role of locoregional therapy and predictors for dropout in patients with hepatocellular carcinoma listed for liver transplantation. *J. Vasc. Interv. Radiol.* 26(12), 1761–1768 (2015).
12. Chaiteerakij R, Addissie BD, Roberts LR. Update on biomarkers of hepatocellular carcinoma. *Clin. Gastroenterol. Hepatol.* 13(2), 237–245 (2015).
13. Halazun KJ, Hardy MA, Rana AA *et al.* Negative impact of neutrophil-lymphocyte ratio on outcome after liver transplantation for hepatocellular carcinoma. *Ann. Surg.* 250(1), 141–151 (2009).
14. Chew V, Chen J, Lee D *et al.* Chemokine-driven lymphocyte infiltration: an early intratumoural event determining long-term survival in resectable hepatocellular carcinoma. *Gut* 61(3), 427–438 (2012).
15. Garnelo M, Tan A, Her Z *et al.* Interaction between tumour-infiltrating B cells and T cells controls the progression of hepatocellular carcinoma. *Gut* 66(2), 342–351 (2017).
16. Unitt E, Marshall A, Gelson W *et al.* Tumour lymphocytic infiltrate and recurrence of hepatocellular carcinoma following liver transplantation. *J. Hepatol.* 45(2), 246–253 (2006).
17. Yang XH, Yamagiwa S, Ichida T *et al.* Increase of $CD4^+$ $CD25^+$ regulatory T-cells in the liver of patients with hepatocellular carcinoma. *J. Hepatol.* 45(2), 254–262 (2006).
18. Dutkowski P, de Rougemont O, Mullhaupt B, Clavien PA. Current and future trends in liver transplantation in Europe. *Gastroenterology* 138(3), 802-809.e801–804 (2010).
19. Washburn K, Edwards E, Harper A, Freeman R. Hepatocellular carcinoma patients are advantaged in the current liver transplant allocation system. *Am. J. Transplant* 10(7), 1643–1648 (2010).
20. Ishaque T, Massie AB, Bowring MG *et al.* Liver transplantation and waitlist mortality for HCC and non-HCC candidates following the 2015 HCC exception policy change. *Am. J. Transplant* 19(2), 564–572 (2019).
21. Mazzaferro V. Squaring the circle of selection and allocation in liver transplantation for HCC: an adaptive approach. *Hepatology* 63(5), 1707–1717 (2016).
22. Costentin CE, Bababekov YJ, Zhu AX, Yeh H. Is it time to reconsider the milan criteria for selecting patients with hepatocellular carcinoma for deceased-donor liver transplantation? *Hepatology* 69(3), 1324–1336 (2019).
23. Dubay D, Sandroussi C, Sandhu L *et al.* Liver transplantation for advanced hepatocellular carcinoma using poor tumor differentiation on biopsy as an exclusion criterion. *Ann. Surg.* 253(1), 166–172 (2011).
24. 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* 143(4), 986–994.e983 (2012).
25. Grat M, Kornasiewicz O, Lewandowski Z *et al.* Combination of morphologic criteria and alpha-fetoprotein in selection of patients with hepatocellular carcinoma for liver transplantation minimizes the problem of posttransplant tumor recurrence. *World J. Surg.* 38(10), 2698–2707 (2014).
26. 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.* 264(5), 787–796 (2016).
27. 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.* 263(5), 842–850 (2016).
28. Marvin MR, Ferguson N, Cannon RM, Jones CM, Brock GN. $MELD_{EQ}$: an alternative model for end-stage liver disease score for patients with hepatocellular carcinoma. *Liver Transpl.* 21(5), 612–622 (2015).
29. Mazzaferro V, Sposito C, Zhou J *et al.* Metroticket 2.0 model for analysis of competing risks of death after liver transplantation for hepatocellular carcinoma. *Gastroenterology* 154(1), 128–139 (2018).

30. 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.* 3(4), 493–500 (2017).
31. Sasaki K, Firl DJ, Hashimoto K *et al.* Development and validation of the HALT-HCC score to predict mortality in liver transplant recipients with hepatocellular carcinoma: a retrospective cohort analysis. *Lancet Gastroenterol. Hepatol.* 2(8), 595–603 (2017).
32. Toso C, Majno P, Berney T, Morel P, Mentha G, Combescure C. Validation of a dropout assessment model of candidates with/without hepatocellular carcinoma on a common liver transplant waiting list. *Transpl. Int.* 27(7), 686–695 (2014).
33. Toso C, Meeberg G, Hernandez-Alejandro R *et al.* Total tumor volume and alpha-fetoprotein for selection of transplant candidates with hepatocellular carcinoma: a prospective validation. *Hepatology* 62(1), 158–165 (2015).
34. Halazun KJ, Najjar M, Abdelmessih RM *et al.* Recurrence after liver transplantation for hepatocellular carcinoma: a new MORAL to the story. *Ann. Surg.* 265(3), 557–564 (2017).
35. Takada Y, Ito T, Ueda M *et al.* Living donor liver transplantation for patients with HCC exceeding the Milan criteria: a proposal of expanded criteria. *Dig. Dis.* 25(4), 299–302 (2007).
36. Zhang B-H, Yang B-H, Tang Z-Y. Randomized controlled trial of screening for hepatocellular carcinoma. *J. Cancer Res. Clin. Oncol.* 130(7), 417–422 (2004).
37. Berry K, Ioannou GN. Serum alpha-fetoprotein level independently predicts posttransplant survival in patients with hepatocellular carcinoma. *Liver Transpl.* 19(6), 634–645 (2013).
38. Giannini EG, Marenco S, Borgonovo G *et al.* Alpha-fetoprotein has no prognostic role in small hepatocellular carcinoma identified during surveillance in compensated cirrhosis. *Hepatology* 56(4), 1371–1379 (2012).
39. Kinoshita A, Onoda H, Imai N *et al.* Comparison of the prognostic value of inflammation-based prognostic scores in patients with hepatocellular carcinoma. *Br. J. Cancer* 107(6), 988–993 (2012).
40. Kalra A, Wedd JP, Bambha KM *et al.* Neutrophil-to-lymphocyte ratio correlates with proinflammatory neutrophils and predicts death in low model for end-stage liver disease patients with cirrhosis. *Liver Transpl.* 23(2), 155–165 (2017).
41. Motomura T, Shirabe K, Mano Y *et al.* Neutrophil-lymphocyte ratio reflects hepatocellular carcinoma recurrence after liver transplantation via inflammatory microenvironment. *J. Hepatol.* 58(1), 58–64 (2013).
42. Mcvey JC, Sasaki K, Firl DJ *et al.* Prognostication of inflammatory cells in liver transplantation: is the waitlist neutrophil-to-lymphocyte ratio really predictive of tumor biology? *Clin. Transplant.* 33(12), e13743 (2019).
43. Fujiki M, Takada Y, Ogura Y *et al.* Significance of des-gamma-carboxy prothrombin in selection criteria for living donor liver transplantation for hepatocellular carcinoma. *Am. J. Transplant.* 9(10), 2362–2371 (2009).
44. Yao FY, Mehta N, Flemming J *et al.* Downstaging of hepatocellular cancer before liver transplant: long-term outcome compared to tumors within Milan criteria. *Hepatology* 61(6), 1968–1977 (2015).
45. Cucchetti A, Serenari M, Sposito C *et al.* Including mRECIST in the Metroticket 2.0 criteria improves prediction of hepatocellular carcinoma-related death after liver transplant. *J. Hepatol.* 73(2), 342–348 (2020).
46. Firl DJ, Kimura S, McVey J *et al.* Reframing the approach to patients with hepatocellular carcinoma: longitudinal assessment with hazard associated with liver transplantation for HCC (HALTHCC) improves ablate and wait strategy. *Hepatology* 68(4), 1448–1458 (2018).
47. Kurtz DM, Esfahani MS, Scherer F *et al.* Dynamic risk profiling using serial tumor biomarkers for personalized outcome prediction. *Cell* 178(3), 699–713.e619 (2019).
48. Vibert E, Schwartz M, Olthoff KM. Advances in resection and transplantation for hepatocellular carcinoma. *J. Hepatol.* 72(2), 262–276 (2020).
49. 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.* 10(1), 35–43 (2009).
50. Sugawara Y, Tamura S, Makuuchi M. Living donor liver transplantation for hepatocellular carcinoma: Tokyo University series. *Dig. Dis.* 25(4), 310–312 (2007).
51. Toso C, Asthana S, Bigam DL, Shapiro AM, Kneteman NM. Reassessing selection criteria prior to liver transplantation for hepatocellular carcinoma utilizing the Scientific Registry of Transplant Recipients database. *Hepatology* 49(3), 832–838 (2009).
52. Alver SK, Lorenz DJ, Washburn K, Marvin MR, Brock GN. Comparison of two equivalent model for end-stage liver disease scores for hepatocellular carcinoma patients using data from the United Network for Organ Sharing liver transplant waiting list registry. *Transpl. Int.* 30(11), 1098–1109 (2017).
53. Decaens T, Roudot-Thoraval F, Hadni-Bresson S *et al.* Impact of UCSF criteria according to pre- and post-OLT tumor features: analysis of 479 patients listed for HCC with a short waiting time. *Liver Transpl.* 12(12), 1761–1769 (2006).
54. Leung JY, Zhu AX, Gordon FD *et al.* Liver transplantation outcomes for early-stage hepatocellular carcinoma: results of a multicenter study. *Liver Transpl.* 10(11), 1343–1354 (2004).
55. Schwartz M. Liver transplantation for hepatocellular carcinoma. *Gastroenterology* 127(1 Suppl. 5), S268–S276 (2004).
56. Yao FY, Xiao L, Bass NM, Kerlan R, Ascher NL, Roberts JP. Liver transplantation for hepatocellular carcinoma: validation of the UCSF-expanded criteria based on preoperative imaging. *Am. J. Transpl.* 7(11), 2587–2596 (2007).

57. D'Amico F, Schwartz M, Vitale A *et al.* Predicting recurrence after liver transplantation in patients with hepatocellular carcinoma exceeding the up-to-seven criteria. *Liver Transpl.* 15(10), 1278–1287 (2009).
58. Lei JY, Wang WT, Yan LN. Up-to-seven criteria for hepatocellular carcinoma liver transplantation: a single center analysis. *World J. Gastroenterol.* 19(36), 6077–6083 (2013).
59. 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.* 66(3), 552–559 (2017).
60. Firl DJ, Sasaki K, Agopian VG *et al.* Charting the path forward for risk prediction in liver transplant for hepatocellular carcinoma: international validation of HALTHCC among 4,089 patients. *Hepatology* 71(2), 569–582 (2019).
61. Durand F, Antoine C, Soubrane O. Liver transplantation in France. *Liver Transpl.* 25(5), 763–770 (2019).
62. Goel A, Kim WR, Pyke J *et al.* Liver simulated allocation modeling: were the predictions accurate for share 35? *Transplantation* 102(5), 769–774 (2018).
63. Thompson D, Waisanen L, Wolfe R, Merion RM, Mccullough K, Rodgers A. Simulating the allocation of organs for transplantation. *Health Care Manag. Sci.* 7(4), 331–338 (2004).
64. Kilambi V, Bui K, Mehrotra S. LivSim: an open-source simulation software platform for community research and development for liver allocation policies. *Transplantation* 102(2), e47–e48 (2018).
65. Linecker M, Krones T, Berg T *et al.* Potentially inappropriate liver transplantation in the era of the 'sickest first' policy – a search for the upper limits. *J. Hepatol.* 68(4), 798–813 (2018).
66. El-Khoueiry AB, Sangro B, Yau T *et al.* Nivolumab in patients with advanced hepatocellular carcinoma (CheckMate 040): an open-label, non-comparative, Phase I/II dose escalation and expansion trial. *Lancet* 389(10088), 2492–2502 (2017).
67. Llovet JM, Ricci S, Mazzaferro V *et al.* Sorafenib in advanced hepatocellular carcinoma. *N. Engl. J. Med.* 359(4), 378–390 (2008).
68. Bruix J, Takayama T, Mazzaferro V *et al.* Adjuvant sorafenib for hepatocellular carcinoma after resection or ablation (STORM): a Phase III, randomised, double-blind, placebo-controlled trial. *Lancet. Oncol.* 16(13), 1344–1354 (2015).
69. Lee JH, Lee J-H, Lim Y-S *et al.* Adjuvant immunotherapy with autologous cytokine-induced killer cells for hepatocellular carcinoma. *Gastroenterology* 148(7), 1383–1391.e1386 (2015).