Elsevier Editorial System(tm) for Journal of Hepatology

Manuscript Draft

Manuscript Number: JHEPAT-D-19-00172

Title: JHEPAT-D-18-01385 "Randomized trials and endpoints in advanced HCC: Role of PFS as a surrogate of survival."

Article Type: Proposal - Editorial

Section/Category: Clinical hepatocellular carcinoma (HCC)

Keywords:

Corresponding Author: ,

Corresponding Author's Institution:

First Author:

Order of Authors:

Elsevier Editorial System(tm) for Journal of Hepatology

Manuscript Draft

Manuscript Number: JHEPAT-D-18-01385R3

Title: Randomized trials and endpoints in advanced HCC: Role of PFS as a surrogate of survival.

Article Type: Invited Review

Section/Category: Clinical hepatocellular carcinoma (HCC)

Keywords: liver cancer; systemic therapies; FDA approval.

Corresponding Author: Dr. Josep M. Llovet, MD

Corresponding Author's Institution: HCC Translational Research Laboratory, BCLC Group, Liver Unit. Institut d'Investigacions Biomediques Agusto Pi I Sunyer (IDIBAPS), Hospital Clinic, Barcelona, 08036, Spain. Division of Liver Diseases, Mount Sinai School of Medicine, New York, NY, USA

First Author: Josep M. Llovet, MD

Order of Authors: Josep M. Llovet, MD; Robert Montal; Augusto Villanueva

Manuscript Region of Origin: SPAIN

Abstract: Hepatocellular carcinoma (HCC) is a relevant cause of cancer-related mortality worldwide. Around half of HCC patients will receive systemic therapies during their life span. The pivotal positive sorafenib trial and regulatory approval in 2007 was followed by a decade of negative studies with drugs leading to marginal anti-tumoral efficacy, toxicity, or trials with lack of enrichment strategies. This trend has changed during the period 2016-18, when several compounds such as lenvatinib (in first line) and regorafenib, cabozantinib, ramucirumab and nivolumab (in second-line) showed clinical benefit. These successes came at a cost of increasing the complexity of decision-making, and ultimately, impacting the design of future clinical trials. Nowadays, life expectancy with single active agents has surpassed the threshold of 1 year and the field is facing encouraging outcomes ~2 years with sequential strategies. Overall survival (OS) remains as the main endpoint in phase III investigations, but as in other solid tumors, there is a clear need to define surrogate endpoints that both reliably recapitulate survival benefits and can be assessed prior additional efficacious drugs are administered. A thorough analysis of 21 phase III trials published in advanced HCC demonstrated a moderate correlation between progression-free survival (PFS) or time to progression (TTP) with OS (R=0.84 and R=0.83, respectively). Nonetheless, significant differences in PFS were only followed by differences in survival in 3 out of 7 phase III studies. In these later cases, the magnitude of benefit for PFS was HR ≤ 0.6, and thus this threshold is herein proposed as a potential surrogate endpoint of OS in advanced HCC. Conversely, PFS with a HR between 0.6-0.7, despite significant, was not associated with better survival, and thus these magnitudes are considered uncertain surrogates. In the current review, we discuss the reasons for positive or negative phase III trials in advanced

HCC, and the strengths and limitations of clinical surrogate endpoints (PFS, TTP and objective response rate [ORR]) to predict survival.

Response to Reviewers: Please see attached file

Josep M Llovet, MD
Professor of Research-ICREA
Translational research in liver cancer
IDIBAPS, Liver Unit, Hospital Clínic
University of Barcelona
Villarroel 170
08036 Barcelona
Catalonia, Spain
Office: +34-932279156
Lab: +34-932279155
Fax: +34-932275792
E-mail: jmllovet@clinic.cat
Assistant: Ariadna Farré (afarrec@clinic.cat)

Barcelona, December 19th, 2018

To: Dr. Rajiv Jalan, Editor-in-Chief, Journal of Hepatology
Dr. Jessica Zucman-Rossi, Co-Editor, Journal of Hepatology
Dr. Jean-Charles Nault, Associate Editor, Journal of Hepatology

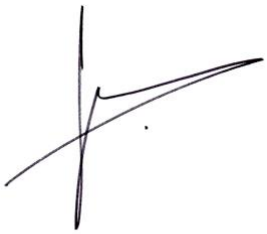
Dear Editors,

Thank you for considering our manuscript JHEPAT-D-18-01385R2, entitled: **"Randomized trials and endpoints in advanced HCC: Role of PFS as a surrogate of survival"** by Llovet JM et al. for further evaluation in order to be published in Journal of Hepatology.

We have modified the manuscript to address some of the reviewers' concerns/suggestions, including additional analyses following their recommendations in order to further strengthen the information obtained with the meta-analysis of phase III clinical trials conducted in advanced HCC.

In the document enclosed, we describe the point-by-point response with changes to the manuscript text underlined to assist reviewers and include a clean final document. We hope that you will find the revised manuscript finally acceptable for publication in your prestigious Journal.

Sincerely,



Josep M Llovet, MD

Josep M Llovet, MD
Professor of Research-ICREA
Translational research in liver cancer
IDIBAPS, Liver Unit, Hospital Clínic
University of Barcelona
Villarroel 170
08036 Barcelona
Catalonia, Spain
Office: +34-932279156
Lab: +34-932279155
Fax: +34-932275792
E-mail: jmllovet@clinic.cat
Assistant: Ariadna Farré (afarrec@clinic.cat)

POINT-BY-POINT RESPONSE (JHEPAT-D-18-01385R2)

Dear Editors and Reviewers,

We appreciate your comments on our manuscript. In the version enclosed we have addressed the points requested by reviewers and are positive that the revised version including your input has improved the overall message of the manuscript. Please, find below our point-by-point responses to the comments. In order to assist reviewers, we have established four writing modes in this document: **black-bold for the original reviewer's comment as included in the decision letter**, black-regular for our responses to reviewer's comments, *italic for sentences that were already present in the manuscript* and *underlined for changes made to the original version.* References already cited in the manuscript or added to the revised manuscript are numbered as per the new version of the manuscript.

Reviewer #1:

Despite being requested at all stages in the review process by both reviewers and the editor, the authors have chosen not to address the key issue of the SHAPRP and AP data. Their exclusion is not even mentioned as a limitation. According the extraordinary correlation between TTP and PFS it can be inferred that the HR for PFS was between 0.5 and 0.7 for both trials yet both trials were positive. So you have 3 trials positive with HR <0.5 and 2 with HR > 0.5. This undermines the reliability of 0.5 threshold even before one considers the many other well-reported confounders of the relationship between PFS and OS.

We thank again Reviewer #1 for this thoughtful observation. We agree that including PFS data for SHARP and AP trials in the meta-analysis would be ideal. However, we do not have access to the individual patient data of these trials [7-8] and thus cannot be included it in the manuscript. Despite the role of Dr. Llovet as co-PI in the SHARP clinical trial, he does not have access to the individual patient data, owned by the sponsor of the trial. Analysis of individual data to

Josep M Llovet, MD
Professor of Research-ICREA
Translational research in liver cancer
IDIBAPS, Liver Unit, Hospital Clínic
University of Barcelona
Villarroel 170
08036 Barcelona
Catalonia, Spain
Office: +34-932279156
Lab: +34-932279155
Fax: +34-932275792
E-mail: jmllovet@clinic.cat
Assistant: Ariadna Farré (afarrec@clinic.cat)

explore the actual PFS of both trails will require an agreement with the sponsor beyond the scope of this review. Nonetheless, we are aware of the importance of the point highlighted by the reviewed and have now tried to address it in an indirect way by providing an estimation of the PFS. In order to estimate PFS in these 2 trials we took advantage of the high trial-level correlation observed between TTP and PFS in advanced HCC (Figure 4). According to the linear equation obtained comparing both surrogate endpoints [log HR_{PFS} = 0.014 + 0.927 x log HR_{TTP}], the inferred PFS HR of SHARP and AP trials are close to 0.60. Therefore, we have amended our 0.5 threshold for PFS, which certainly can be too stringent and have modified the proposed minimum threshold effect of PFS to 0.6. With this change, we are decreasing the positive predictive value of PFS to predict significant differences in OS, but certainly are capturing the trials for which PFS will be borderline and are certainly associated with a magnitude of clinical benefit in terms of OS. Slightly beyond that threshold of =0.6, we have examples of RCT that were positive for PFS but negative for OS (REACH HR-PFS=0.63 and HR-OS=0.87 and EACH HR-PFS=0.62; HR-OS=0.80). As a result of these changes we have modified the threshold in Fig 3, incorporating the estimated PFS of SHARP and Asian – Pacific trials.

During this period of time, two recent studies have been released supporting our threshold of HR ≤ 0.6 for PFS, and thus we have included in the current submission. The first one, an individual-patient data meta-analysis of two randomized clinical trials (REACH[21] and REACH-2[25]), showing a significant OS HR with a PFS HR of 0.57 (Llovet et al, Hepatology 2018; 68, AASLD 2018 meeting abstract). The second one, a phase 3 randomized clinical trial comparing sorafenib with or without cTACE in advanced HCC, a negative study for its primary endpoint (OS) with a PFS HR of 0.73 (JW Park et al, Journal of Hepatology 2018; in Press).

Josep M Llovet, MD
Professor of Research-ICREA
Translational research in liver cancer
IDIBAPS, Liver Unit, Hospital Clínic
University of Barcelona
Villarroel 170
08036 Barcelona
Catalonia, Spain
Office: +34-932279156
Lab: +34-932279155
Fax: +34-932275792
E-mail: jmllovet@clinic.cat
Assistant: Ariadna Farré (afarrec@clinic.cat)

Figure 3

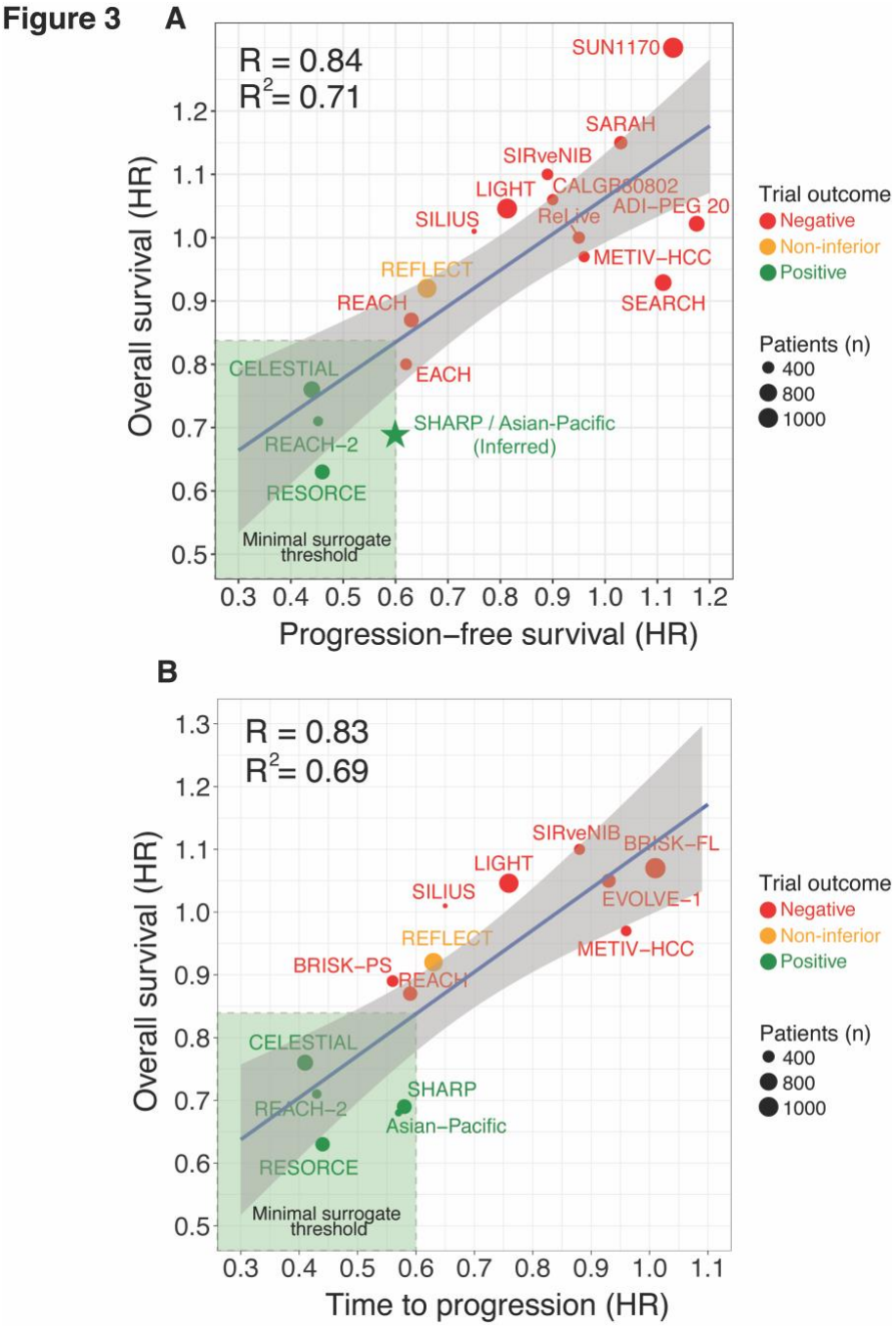


Figure 3: ***Correlation between surrogate endpoints (PFS[A] and TTP[B]) and hard endpoint (OS).***

Trial-level correlation between endpoints using criteria from the Institute for Quality and Efficiency in Health Care (IQWIG). R and R2 refers to the weighted Pearson coefficient between the HR of OS and the HR of the surrogate endpoint. IQWIG categorizes the strength of the correlation based on the value of R as low (R<0.7), moderate (R>0.7

Josep M Llovet, MD
Professor of Research-ICREA
Translational research in liver cancer
IDIBAPS, Liver Unit, Hospital Clínic
University of Barcelona
Villarroel 170
08036 Barcelona
Catalonia, Spain
Office: +34-932279156
Lab: +34-932279155
Fax: +34-932275792
E-mail: jmllovet@clinic.cat
Assistant: Ariadna Farré (afarrec@clinic.cat)

to R<0.85) and high (R>0.85)[85]. Each dot represents one of the phase III clinical trials conducted on advanced HCC. Size of the dot is proportional to the total number of patients enrolled in the trial. Trials are colored based on whether the final result was positive, negative or non-inferior. X and Y axis depict the value of the HR for the surrogate (TTP or PFS) and the hard endpoint (OS), respectively.

In summary, the changes in the text have been as follows:

[Page 2, line 28]: *In these later cases, the magnitude of benefit for PFS was HR ≤ 0.6, and thus this threshold is herein proposed as a potential surrogate endpoint of OS in advanced HCC. Conversely, PFS with a HR between 0.6-0.7, despite significant, was not associated with better survival, and thus these magnitudes are considered uncertain surrogates.*

[Page 4, line 38]: *Based on this analysis and a conservative approach to define surrogates of OS in HCC, we propose PFS with a threshold of HR≤ 0.6 as reliable surrogate with solid positive predictive value, whereas the threshold of HR=0.6-0.7 -despite leading to positive statistical results- is defined as clinically uncertain in terms of capturing true advantages in OS.*

[Page 12, line 5]: *When specifically analyzing the positive predictive value of theoretical thresholds of PFS correlating with OS, only 3/7 PFS reported a HR ≤0.6 that was significantly associated with a positive survival clinical benefit (in all cases with a HR for OS < 0.8). Conversely, those four studies reporting a positive PFS with a HR between 0.6-0.8 were associated with no significant survival benefits (HR for OS between 0.87 and 1.05) (Figure 3A). In our study, according to the linear regression equation obtained [log HR_{OS} = 0.072 + 0.487 x log HR_{PFS}], a threshold of PFS HR=0.6 (representing a 40% risk reduction) will decrease ~17% the risk of OS (OS HR=0.83) (see Figure 3A). In summary, a moderate correlation has been established between PFS and OS in 21 RCTS in advanced HCC. A value of HR ≤0.6 is proposed as surrogate threshold effect [89], and is likely to predict a clinically meaningful improvement in OS.*

Josep M Llovet, MD
Professor of Research-ICREA
Translational research in liver cancer
IDIBAPS, Liver Unit, Hospital Clínic
University of Barcelona
Villarroel 170
08036 Barcelona
Catalonia, Spain
Office: +34-932279156
Lab: +34-932279155
Fax: +34-932275792
E-mail: jmllovet@clinic.cat
Assistant: Ariadna Farré (afarrec@clinic.cat)

[Page 13, line 13]: *In fact, when we inferred the non-reported PFS HR of SHARP and AP trials according to the linear equation obtained comparing both surrogate endpoints [log HR_{PFS} = 0.014 + 0.927 x log HR_{TTP}], the HR values are close to 0.60, just at the previously proposed minimum threshold.*

[Page 15, line 48]: *In this scenario, PFS has shown moderate correlation with OS (R=0.84), and a threshold of HR≤0.6 defines a conservative approach of surrogate endpoint able to capture survival differences in a superiority trial with a high positive predictive value. Two recent studies have been released supporting our threshold of HR ≤ 0.6 for PFS. The first one, an individual-patient data meta-analysis of two RCTs (REACH[21] and REACH-2[25], showing a significant OS HR with a PFS HR of 0.57[129]. The second one, a phase 3 RCT comparing sorafenib with or without cTACE in advanced HCC, a negative study for its primary endpoint (OS) with a PFS HR of 0.73[130]. Thus, PFS-HR ≤0.6 can be considered a candidate endpoint in phase II and phase III RCTs when subsequent therapies are expected to impact overall outcome.*

[Page 16, line 24]: *PFS has a moderate correlation at trial level with OS (R=0.84). A conservative minimum surrogate threshold effect of HR ≤0.6 is highly predictive of a significant improvement in OS, whereas HR ranging from 0.6 – 0.7 are uncertain surrogates.*

[129] Llovet JM, Kudo M, Finn R, Galle PR, Blanc J, Okusaka T, et al. Ramucirumab As Second-Line Treatment in Patients with Hepatocellular Carcinoma (HCC) and Elevated Alpha-Fetoprotein (AFP) Following Sorafenib: Pooled Results from Two Global Phase 3 Studies (REACH-2 and REACH). Hepatology 2018;68, suppl.

[130]Park J-W, Kim YJ, Kim DY, Bae SH, Paik SW, Lee Y-J. Sorafenib with or without concurrent transarterial chemoembolization in patients with advanced hepatocellular carcinoma: a phase III STAH trial. J Hepatol 2018;In Press.

Josep M Llovet, MD
Professor of Research-ICREA
Translational research in liver cancer
IDIBAPS, Liver Unit, Hospital Clínic
University of Barcelona
Villarroel 170
08036 Barcelona
Catalonia, Spain
Office: +34-932279156
Lab: +34-932279155
Fax: +34-932275792
E-mail: jmllovet@clinic.cat
Assistant: Ariadna Farré (afarrec@clinic.cat)

Reviewer #3:

The revised version has improved the overall message of the manuscript although there is still a tendency, as the first reviewer notes, for the authors to placate the reviewer by agreeing with their criticisms and then to make no related charges in the text.

[Trial data belongs to the company sponsoring the study. Further exploring so is far beyond the scope of this review]

Understood, But the question is, do the authors have access to the raw data on which SHARP and the AP studies were based? Are they in a position to undertake their own independent analyses? Or have the trial results and all the subsequent downstream analyses been undertaken by the company?

We have made some changes addressing this issue. Please see response to Reviewer #1.

[To answer the question about the comparison between different radiological criteria, this concept should be addressed in an independent meta-analysis with trials using RECIST or mRECIST]

Agreed. Then since radiological criteria of response are central to this review, the author's stance should surely be, words to the effect that, the optimal method of assessing response has not yet been defined and awaits an independent meta-analysis. Then, by all means continue 'nonetheless recent studies…suggest'. But if this is done the caveats described by Bruix et al (Journal of Hepatology 2017 vol. 66 1114-1117) should be noted for balance.

As it stands p13 &14 give the impression that mRECIST is clearly superior.

Josep M Llovet, MD
Professor of Research-ICREA
Translational research in liver cancer
IDIBAPS, Liver Unit, Hospital Clínic
University of Barcelona
Villarroel 170
08036 Barcelona
Catalonia, Spain
Office: +34-932279156
Lab: +34-932279155
Fax: +34-932275792
E-mail: jmllovet@clinic.cat
Assistant: Ariadna Farré (afarrec@clinic.cat)

We have modified this section of the text to better reflect the controversy regarding methods to evaluate tumor response in HCC.

[Page 14, line 24]: *A direct comparison between RECIST and mRECIST for OS surrogacy through an independent meta-analysis of trials using either criteria would be ideal to define the role of ORR to predict OS in advanced HCC. This will also help determine the best tool to evaluate tumor response to systemic therapies. However, since only 5 RCTs reported response data using mRECIST, we did not sub-analyze this endpoint according to the tool used to evaluate response.*

[In fact, it is not intended that these RECIST guidelines play a role in the decision making clinical process]

Is it not proposed that patients who progress on Sorafenib, according to RECIST guidelines, should be moved from Sorafenib to Regorafenib?

The scope of this review is oriented to discuss clinical trial design and recent developments based on the recent trials reported in HCC. The current EASL guidelines recommend both RECIST and mRECIST to assess response to systemic therapies in HCC, so it is difficult to recommend one versus the other with the current data available. Their value is mostly in the context of clinical trials, and that's the main rationale for their initial development (our comment was in this direction).

[The signal was not based on response rate, but on median survival time of the patients receiving sorafenib]

Josep M Llovet, MD
Professor of Research-ICREA
Translational research in liver cancer
IDIBAPS, Liver Unit, Hospital Clínic
University of Barcelona
Villarroel 170
08036 Barcelona
Catalonia, Spain
Office: +34-932279156
Lab: +34-932279155
Fax: +34-932275792
E-mail: jmllovet@clinic.cat
Assistant: Ariadna Farré (afarrec@clinic.cat)

Understood and point taken.

[Positive and negative trials are defined according to the primary endpoint of the study] Understood. Then this should be stated, it is not self-evident.

Table 1 summarizes the criteria used to define positive (for superiority and non-inferiority) and negative trails. The primary endpoint of all phase III trials included is OS.

[We agree this comment. The present review focus mainly on RCT but we thought that the recent success of checkpoint inhibitors in HCC should be introduced. It is clearly stated throughout the manuscript that the accelerated approval of nivolumab is based on a phase II single-arm trial]

The reviewers point was merely that the heading of this section is phase III trials - so phase II studies seem inappropriate in this section.

We have modified the heading of the section to better reflect this:

Overview of phase III and practice-changing phase II trials reported during the last 10 years

[If we understand well, the reviewer suggests incorporating additional data regarding reference 34, beyond that sorafenib is more effective in HCV-related HCC]

Yes, if a strength of the SHARP trial is that it was stratified for extra-hepatic disease (EHD), then doesn't' figure 1 (A & B) in ref 34 (and the associated tables [4]) become an important

Josep M Llovet, MD
Professor of Research-ICREA
Translational research in liver cancer
IDIBAPS, Liver Unit, Hospital Clínic
University of Barcelona
Villarroel 170
08036 Barcelona
Catalonia, Spain
Office: +34-932279156
Lab: +34-932279155
Fax: +34-932275792
E-mail: jmllovet@clinic.cat
Assistant: Ariadna Farré (afarrec@clinic.cat)

finding? Shouldn't the fact that in this analysis there is no obvious benefit for sorafenib in EHD patients be noted/commented on in a review of 'randomised trials in HCC?'

The findings observed in the individual-patient data analysis of SHARP and Asian-Pacific trials are relevant enough, as mentioned by Reviewer #3, to be included in the present review:

[Page 4, line 29]: *Etiology is not considered a prognostic factor, but needs to be incorporated when testing sorafenib, since it has been demonstrated to be a predictor of response for this drug, at the same level than absence of extrahepatic spread and low neutrophil-to-lymphocyte ratio[34].*

[Well, this is the whole purpose of a review, to provide a summary of information but also opinions from the authors. I understand that this issue might be debatable, and we are open to this debate]

Well, no.

The purpose of a review may be to express the opinions of the authors, but the opinions should be based on evidence found in a review of the literature. As previously pointed out, the authors offer an unreferenced opinion related to an unwritten law from unidentified experts. As the authors know well expert opinion is not regarded highly. This really matters. Future trials will merely say 'our survival improvement of 2 months is considered clinically significant' and reference this review. There will be no understanding by the reader that there is no evidence on which this statement is based.

The alternative is to briefly review the extensive literature on assessment of clinical benefit. The ESMO (and/or ASCO) guidelines are a good point to start - ESMO-Magnitude of Clinical Benefit Scale version Annals of Oncology 28: 2340-2366, 2017.

Josep M Llovet, MD
Professor of Research-ICREA
Translational research in liver cancer
IDIBAPS, Liver Unit, Hospital Clínic
University of Barcelona
Villarroel 170
08036 Barcelona
Catalonia, Spain
Office: +34-932279156
Lab: +34-932279155
Fax: +34-932275792
E-mail: jmllovet@clinic.cat
Assistant: Ariadna Farré (afarrec@clinic.cat)

The review summarizes the views of the authors based on the current available data, and we agree that this corresponds to limited level of evidence. Our intention is not to overinterpret the data, but to provide a balance view of a key issue that is emerging as a result of the recent changes in the systemic therapy arena in HCC. We agree that the definition of clinical benefit is controversial, highly dependent on the expected outcome for the target population, with confronted opinions between patients, providers, regulatory agencies and health insurers. Accordingly, we have deleted the sentence in which 2 months is considered to be clinically relevant. In addition, we have included the ESMO guidelines that review this topic in detail. Also, we have considered the threshold proposed by ASCO of OS with HR<0.8 as cliinically sound in HCC for magnitude of benefit. (Ellis et al, JCO 2014).

[Page 10, line 10]: *What is the magnitude of benefit to define it as clinically meaningful? This is a controversial topic, highly dependent on the expected outcome for the target population, with confronted opinions between patients, providers, regulatory agencies and health insurers [81,82]. In HCC, there is no consensus on what absolute survival benefit (or the magnitude of benefit in OS as per HR) to be defined as clinically relevant. Reported thresholds of OS with HR<0.8 are also sound for advanced HCC trials[83].* ~~*An unwritten rule among experts estimate that in advanced HCC scenario, where natural outcome (placebo arm) is estimated to be of around 8 months, absolute gains beyond 2 months are considered clinically relevant, while those below 1 months are not.*~~

[82] Cherny NI, Dafni U, Bogaerts J, Latino NJ, Pentheroudakis G, Douillard JY, et al. ESMO-Magnitude of Clinical Benefit Scale version 1.1. Ann Oncol 2017;28:2340–66.

[83] Ellis LM, Bernstein DS, Voest EE, Berlin JD, Sargent D, Cortazar P, et al. American society of clinical oncology perspective: Raising the bar for clinical trials by defining clinically meaningful outcomes. J Clin Oncol 2014;32:1277–80.

[A conservative minimum surrogate threshold effect of HR =0.5 is highly predictive of a significant improvement in OS, whereas HR ranging from 0.5 - 0.7 are uncertain surrogates]

Josep M Llovet, MD
Professor of Research-ICREA
Translational research in liver cancer
IDIBAPS, Liver Unit, Hospital Clínic
University of Barcelona
Villarroel 170
08036 Barcelona
Catalonia, Spain
Office: +34-932279156
Lab: +34-932279155
Fax: +34-932275792
E-mail: jmllovet@clinic.cat
Assistant: Ariadna Farré (afarrec@clinic.cat)

As the authors point out this conclusion is only really supported in the second line setting and then against placebo.

Agreed, see response to Reviewer #1.

[To facilitate the implementation of biomarker-based clinical trials in HCC, it is essential to enforce mandatory tissue collection in all clinical trials testing new compounds[4]]

If biopsy is now indicated in clinical trials then surely:

a) The issue surrounding intra and inter-tumoural heterogeneity needs to be mentioned

b) 'essential to enforce mandatory tissue collection' Where is this in reference 4?

a) The impact of intratumor heterogeneity is still debated as it impacts single-biopsy predictions. We have analyzed the distribution of know driver genes in HCC and it seems that all regions harbor this type of alterations (e.g., TP53, CTNNB1, TERT promoter) (Torrecilla et al. J. Hepatol. 2017, 67, 1222–1231). Other studies have reported similar results (Zhai et al. Nat Commun. 2017 Feb 27;8:4565; Lin et al. Cancer Res. 2017 May 1;77(9):2255-2265). In other tumors, molecular heterogeneity between primary and metastatic tumors is minimal (Reiter et al. Science. 2018 Sep 7;361(6406):1033-1037). We have included a sentence to describe this issue:

[Page 9, line 5]: *To facilitate the implementation of biomarker-based clinical trials in HCC, it is essential to enforce mandatory tissue collection in all clinical trials testing new compounds[4]. In this regard, the impact of intratumor heterogeneity in single-biopsy predictions is still debated despite recent studies have shown that driver gene mutations are common between different regions of the tumor [76,77].*

[76] *Torrecilla S, Sia D, Harrington AN, Zhang Z, Cabellos L, Cornella H, et al. Trunk mutational events present minimal intra- and inter-tumoral heterogeneity in hepatocellular carcinoma. J Hepatol 2017;67:1222–31.*

Josep M Llovet, MD
Professor of Research-ICREA
Translational research in liver cancer
IDIBAPS, Liver Unit, Hospital Clínic
University of Barcelona
Villarroel 170
08036 Barcelona
Catalonia, Spain
Office: +34-932279156
Lab: +34-932279155
Fax: +34-932275792
E-mail: jmllovet@clinic.cat
Assistant: Ariadna Farré (afarrec@clinic.cat)

[77] *Reiter JG, Makohon-Moore AP, Gerold JM, Heyde A, Attiyeh MA, Kohutek ZA, et al. Minimal functional driver gene heterogeneity among untreated metastases. Science 2018;361:1033–7.*

b) One of the EASL HCC guidelines recommendations for trial design and endpoints is:

- Upfront liver biopsy and blood sampling is recommended for clinical and diagnostic trials (recommendation strong).

[Is worth to mention though that this rule is supported mainly by positive trials in the second-line setting comparing active drugs vs placebo. We assume that such association is retained in front-line comparing two active drugs, but recommendation in that setting should be tempered due to the lack of confirmatory data]

Agreed, and the authors seem, in this statement, to undermine the main thesis of this review.

They are making a major assumption when all three trials with HR <0.5 are indeed second line against placebo and, as they acknowledge, there is 'lack of confirmatory data to support this assumption' [in other scenarios]. Yet in the conclusion we find 'Thus, PFS-HR =0.5 can be considered a candidate endpoint in phase II and phase III RCTs when subsequent therapies are expected to impact overall outcome'.

Agree, please see response to Reviewer #1. We mention the potential discrepancy between results in first and second line to provide a balance view of our interpretation of the data and underscore its potential limitations.

Their language must be much more 'tempered' in the conclusions.

Josep M Llovet, MD
Professor of Research-ICREA
Translational research in liver cancer
IDIBAPS, Liver Unit, Hospital Clínic
University of Barcelona
Villarroel 170
08036 Barcelona
Catalonia, Spain
Office: +34-932279156
Lab: +34-932279155
Fax: +34-932275792
E-mail: jmllovet@clinic.cat
Assistant: Ariadna Farré (afarrec@clinic.cat)

As suggested by Reviewer #3 we have tempered the statement that suggests PFS-HR ≤0.6 as a candidate endpoint in advanced HCC:

[Page 15, line 58]: *Thus, PFS-HR ≤0.6 could be considered a candidate endpoint in phase II and phase III RCTs when subsequent therapies are expected to impact overall outcome.*

Reviewer #4:

1. The structure of the paper does not identify clearly a Methods section where we would learn exactly how was the study conducted and which analyses were done precisely. For instance, the end of introduction already gives results. This is perhaps because it is an invited review, but since it aims at providing scientifically grounded recommendations, it should follow the standards for scientific reporting, so that we are able to evaluate the content.

We thank Reviewer #4 for this relevant comment. However, as pointed out, this is not an original article and a methods section is not allowed. We thought that the analyses conducted increase the interest of the review and provide new tools for the successful development of clinical trials in advanced HCC. In the manuscript and figure legends, we have defined the method used for the development of the meta-analysis of trial-level surrogate endpoints:

[Page 11, line 25]: *The Institute for Quality and Efficiency in Health Care has proposed a set of criteria to quantify the association between a surrogate and hard endpoint, which includes low (R<0.7), moderate (R>0.7 to R<0.85) and high correlation (R>0.85)[89]. R refers to the weighted Pearson coefficient between HR of OS and HR of the surrogate endpoint.*

Josep M Llovet, MD
Professor of Research-ICREA
Translational research in liver cancer
IDIBAPS, Liver Unit, Hospital Clínic
University of Barcelona
Villarroel 170
08036 Barcelona
Catalonia, Spain
Office: +34-932279156
Lab: +34-932279155
Fax: +34-932275792
E-mail: jmllovet@clinic.cat
Assistant: Ariadna Farré (afarrec@clinic.cat)

2. Clearly, the selection of the trials for analysis is crucial for the conclusions. I have no idea whether SHARP and AP trials should be part of this review or not, but the trials included should be according to clearly predefined eligibility criteria. The type of drug can be part of the eligibility criteria, the important point being to be clear upon that. It is also important to perform a thorough literature search, and to include all trials according to the eligibility criteria, especially if no individual patient data are used.

The data of SHARP and AP trials are included for the TTP but not for the PFS correlation since these studies did not report PFS. Previous reviewers have asked for this information, but we do not have access to individual patient data to calculate PFS in these trials (see comments to Reviewer #1 for details). As recommended, we have clarified the criteria used to select the 21 trials included in the analysis:

[Page 11, line 32]: *In order to explore the concept of surrogate endpoints recapitulating OS in advanced HCC we have identified 21 RCTs assessing systemic therapies with or without loco-regional therapies in advanced HCC (12 in first-line and 9 in second-line) (Table 1) published between 2008 and 2018 through a MEDLINE search via PubMed using the keywords "advanced hepatocellular carcinoma". Results were limited to "clinical trial, phase III". Trials recently presented at international meetings (2016-2018) were also included despite the full manuscript is not yet available. For each trial, data on sample size, radiological response, TTP and OS were collected. TTP and OS were determined in terms of hazard ratio (HR) using published data (values less than 1 denotes a favourable result in the experimental group). In addition, ORR was established with odds ratio calculated from the published radiological response (values greater than 1 denotes a favourable result in the experimental group).*

3. Validation if surrogate outcomes can involve patient-level and trial-level analyses. At the patient level, the question is e.g. what is the association between a given PFS and OS for a particular patient. At the trial level, the question is rather what is the association

Josep M Llovet, MD
Professor of Research-ICREA
Translational research in liver cancer
IDIBAPS, Liver Unit, Hospital Clínic
University of Barcelona
Villarroel 170
08036 Barcelona
Catalonia, Spain
Office: +34-932279156
Lab: +34-932279155
Fax: +34-932275792
E-mail: jmllovet@clinic.cat
Assistant: Ariadna Farré (afarrec@clinic.cat)

between a given amount of effect on PFS on the treatment effect on OS. Both are very different. Only individual patients data allow to study the first type of surrogacy. Here my view is that only the second one was studied—which I find the most interesting by far—but this should be clarified, and the wording should be adapted for that aim.

We agree with Reviewer #4 about the two types of surrogacy: patient-level and trial-level. That was already acknowledged in the manuscript:

[Page 11, line 20]: *Validation of surrogate endpoints can be conducted at the individual- or trial-levels[88]. While validation of individual-level surrogacy requires individual patient data from at least one clinical trial, trial-level surrogacy uses assembled data from multiple trials.*

In order to stress the type of the analysis conducted in the present review, we have added 'trial-level' in the sentence describing the methodology:

[Page 11, line 43]: *For the purpose of the trial-level analysis, we first assessed the overall correlation between PFS and OS (R=0.84; R^2=0.71) (Figure 3A), and then the correlation of TTP and OS (R=0.83; R^2=0.69) (Figure 3B).*

4. Methods for surrogate outcome validation are complex. The reference 80 is absolutely correct, but there are also other references that should be used to describe (and determine) the methods. For instance papers by Ciani and colleagues in the Journal of Clinical Epidemiology and the BMJ. Then these methods should be described. For instance I do not know whether the gray shaded areas around the regression lines on the surrogacy plots represent confidence intervals or prediction intervals (and we need the latter here).

Josep M Llovet, MD
Professor of Research-ICREA
Translational research in liver cancer
IDIBAPS, Liver Unit, Hospital Clínic
University of Barcelona
Villarroel 170
08036 Barcelona
Catalonia, Spain
Office: +34-932279156
Lab: +34-932279155
Fax: +34-932275792
E-mail: jmllovet@clinic.cat
Assistant: Ariadna Farré (afarrec@clinic.cat)

We agree that there are different methods to study surrogacy in clinical trials. Indeed, the different methods vary in terms of definition of endpoints, measurement of differences between treatments and statistical tools to grade the correlation observed. The revised version mentions that there are alternative methods to conduct this analysis as reported in advanced colorectal cancer by Ciani et al:

[Page 11, line 28]: *R refers to the weighted Pearson coefficient between HR of OS and HR of the surrogate endpoint. Alternative methods to study this correlation have been reported[90].*

[90] Ciani O, Buyse M, Garside R, Peters J, Saad ED, Stein K, et al. Meta-analyses of randomized controlled trials show suboptimal validity of surrogate outcomes for overall survival in advanced colorectal cancer. J Clin Epidemiol 2015;68:833–42.

On the other hand, we have added in figure legends the meaning of the gray shaded areas around the regression lines on the surrogacy plots:

Gray shaded areas represent the upper and lower limits of the 95% confidence intervals for the regression.

5. TTP is a complex outcome. The manuscript mentions the problem of competing risks for PFS, but actually, from a statistical point-of-view, there is no such issue. The issue is more that death can be due to various reasons, some being unrelated to the underlying cancer or treatment effect. But this is part of the outcome definition, as OS for instance. On the contrary, TTP necessitates the handling of competing risks. So it may be relevant

Josep M Llovet, MD
Professor of Research-ICREA
Translational research in liver cancer
IDIBAPS, Liver Unit, Hospital Clínic
University of Barcelona
Villarroel 170
08036 Barcelona
Catalonia, Spain
Office: +34-932279156
Lab: +34-932279155
Fax: +34-932275792
E-mail: jmllovet@clinic.cat
Assistant: Ariadna Farré (afarrec@clinic.cat)

to mention it, and to explain how data have been analyzed. There are several ways (basically two) to handle competing risks. Either cause-specific hazard ratios are computed (this works as if deaths were censored, so I assume this is what the authors refer to), or subdistribution hazard ratios. Both being possibly different, they deserve separate validation for surrogacy. So this should be made explicit.

We thank the reviewer for bringing up the issue of competing risks. In the randomized clinical trials assessed for TTP, deaths were censored as non-radiological progressions at the time of death or at an earlier visit, with a cause-specific hazard, meaning that the risk set decreases each time there is a death. We have clarified this in the revised manuscript:

[Page 12, line 37]: *Deaths are censored as non-radiological progressions at the time of death or at an earlier visit, with a cause-specific hazard, representing informative censoring.*

6. On a slightly similar issue, at least a subgroup analysis for RECIST and mRECIST should be used. These being different response criteria, it should be useful to see whether they perform similarly or not as surrogates.

This issue was already raised in the first point by point review in which Reviewer #4 was not yet included. We did consider conducting this sub-analysis for the manuscript, but we decided not to include it since mRECIST was only reported in 5 HCC randomized trials [10,13,18,19,22]. Thus, we believe that this sub-analysis would be severely underpowered and potentially misleading. We have included a clarification to highlight this limitation:

[Page 14, line 24]: *A direct comparison between RECIST and mRECIST for OS surrogacy through an independent meta-analysis of trials using either criteria would be ideal to define the*

Josep M Llovet, MD
Professor of Research-ICREA
Translational research in liver cancer
IDIBAPS, Liver Unit, Hospital Clínic
University of Barcelona
Villarroel 170
08036 Barcelona
Catalonia, Spain
Office: +34-932279156
Lab: +34-932279155
Fax: +34-932275792
E-mail: jmllovet@clinic.cat
Assistant: Ariadna Farré (afarrec@clinic.cat)

role of ORR to predict OS in advanced HCC. This will also help determine the best tool to evaluate tumor response to systemic therapies. However, since only 5 RCTs reported response data using mRECIST, we did not sub-analyze this endpoint according to the tool used to evaluate response.

7. According to usual definitions (and I'm unsure they were used here) a surrogacy threshold of 0.5 indicates poor surrogacy. So the main conclusions for me is that PFS should rather not be used as a surrogate here.

Please see response to Reviewer #1. The value of 0.5 to define poor surrogacy refers to the correlation between endpoints. The correlation between PFS and OS in HCC is 0.84 (see page 12 of the revised manuscript) which indicates a moderate correlation. We discuss the use of a more stringent value of the HR for PFS (below 0.6) to increase its accuracy to predict OS.

8. I did not see the point of assessing the surrogacy of TTP for PFS. PFS is already a surrogate outcome by itself, and this is of no use if TTP is also assessed for surrogacy to OS.

We agree with Reviewer #4 in this point. However, the reason to include the correlation between TTP and PFS was a specific request from the Editorial Team of the journal in previous revisions of the paper. However, we recognize that Figure 4 is still informative since it shows a strong correlation between both endpoints. In the modern era of HCC trial design, this suggests minimal cirrhosis-related deaths due to restricting patient inclusion to those with well-preserved liver function (Child-Pugh A).

Less major points are:

Josep M Llovet, MD
Professor of Research-ICREA
Translational research in liver cancer
IDIBAPS, Liver Unit, Hospital Clínic
University of Barcelona
Villarroel 170
08036 Barcelona
Catalonia, Spain
Office: +34-932279156
Lab: +34-932279155
Fax: +34-932275792
E-mail: jmllovet@clinic.cat
Assistant: Ariadna Farré (afarrec@clinic.cat)

1. It is unclear in the article how studies are categorized as positive or negative on the plots. Is it the primary outcome of the trial? Or always OS? (could be the same, of course). It would be interesting to classify the points as significant effect on PFS and OS, significant for PFS-NS for OS, NS for PFS-significant for OS, NS for both.

We have added in figure legends how studies are categorized:

Trials are colored based on whether the final result was positive, negative or non-inferior for the primary endpoint (OS).

In Table 1 readers can easily identify those trials where the active arm induced significant differences in one endpoint but not in the others.

2. In the table, HRs should be given with confidence intervals.

As suggested, we have added confidence intervals in Table 1.

Josep M Llovet, MD
Professor of Research-ICREA
Translational research in liver cancer
IDIBAPS, Liver Unit, Hospital Clínic
University of Barcelona
Villarroel 170
08036 Barcelona
Catalonia, Spain
Office: +34-932279156
Lab: +34-932279155
Fax: +34-932275792
E-mail: jmllovet@clinic.cat
Assistant: Ariadna Farré (afarrec@clinic.cat)

	Trial	Arms	N	ORR	TTP		PFS		OS	
					Median	HR	Median	HR	Median	HR
First-line	SHARP (7)	Sorafenib	299	2.3	5.5	0.58 (0.45 - 0.74)	NR		10.7	0.69 (0.55 - 0.87)
		Placebo	303	0.7	2.8				7.9	
	Asian-Pacific (8)	Sorafenib	150	3.3	2.8	0.57 (0.42 - 0.79)	NR		6.5	0.68 (0.50 - 0.93)
		Placebo	76	1.3	1.4				4.2	
	SUN1170 (9)	Sunitinib	530	6.6	4.1	1.13 (0.98 - 1.31)	3.6	1.13 (0.99 - 1.30)	7.9	1.30 (1.13 - 1.50)
		Sorafenib	544	6.1	3.8		3		10.2	
	BRISK-FL* (10)	Brivanib	577	12.0	4.2	1.01 (0.88 - 1.16)	NR		9.5	1.07 (0.94 - 1.23)
		Sorafenib	578	8.8	4.1				9.9	
	LIGHT (11)	Linifanib	514	10.1	5.4	0.76 (0.64 - 0.90)	4.2	0.81 (0.70 - 0.95)	9.1	1.05 (0.90 - 1.22)
		Sorafenib	521	6.1	4		2.9		9.8	
	SEARCH (12)	Sorafenib+Erlotinib	362	6.6	3.2	1.14 (0.94 - 1.37)	NR	1.11 (0.94 - 1.31)	9.5	0.93 (0.78 - 1.11)
		Sorafenib	358	3.9	4				8.5	
	REFLECT* (13)	Lenvatinib	478	24.1	8.9	0.63 (0.53 - 0.73)	7.4	0.66 (0.57 - 0.77)	13.6	0.92 (0.79 - 1.06)
		Sorafenib	476	9.2	3.7		3.7		12.3	
	SARAH (14)	Y90	237	15.2	NR		4.1	1.03 (0.85 - 1.25)	8	1.15 (0.94 - 1.41)
		Sorafenib	222	10.4			3.7		9.9	
	SIRveNIB (15)	Y90	182	16.5	6.1	0.88 (0.7 - 1.1)	5.8	0.89 (0.70 - 1.10)	8.8	1.10 (0.90 - 1.40)
		Sorafenib	178	1.7	5.4		5.1		10	
	EACH (16)	FOLFOX4	184	8.2	NR		2.93	0.62 (0.49 - 0.79)	6.4	0.80 (0.63 - 1.02)
		Doxorubicin	187	2.7			1.77		4.97	
	CALGB80802 (17)	Sorafenib+Doxorubicin	173	NR	NR		3.6	0.90 (0.72 - 1.20)	9.3	1.06 (0.80 - 1.40)
		Sorafenib	173	NR			3.2		10.5	
	SILIUS* (18)	Sorafenib+HAIC	103	36.3	5.3	0.65 (0.48 - 0.87)	4.8	0.75 (0.57 - 1.00)	11.8	1.01 (0.74 - 1.37)
		Sorafenib	103	17.5	3.5		3.5		11.5	
Second-line	BRISK-PS* (19)	Brivanib	263	9.9	4.2	0.56 (0.42 - 0.76)	NR		9.4	0.89 (0.69 - 1.15)
		Placebo	132	1.5	2.7				8.2	
	EVOLVE-1 (20)	Everolimus	362	2.2	3	0.93 (0.75 - 1.15)	NR		7.6	1.05 (0.86 - 1.27)
		Placebo	184	1.6	2.6				7.3	
	REACH (21)	Ramucirumab	283	7.1	3.5	0.59 (0.49 - 0.72)	2.8	0.63 (0.52 - 0.75)	9.2	0.87 (0.72 - 1.05)
		Placebo	282	0.7	2.6		2.1		7.6	
	RESORCE* (22)	Regorafenib	379	10.6	3.2	0.44 (0.36 - 0.55)	3.1	0.46 (0.37 - 0.56)	10.6	0.63 (0.50 - 0.79)
		Placebo	194	4.1	1.5		1.5		7.8	
	METIV-HCC (23)	Tivantinib	226	0.0	2.4	0.96 (0.74 - 1.25)	2.1	0.96 (0.75 - 1.22)	8.4	0.97 (0.75 - 1.25)
		Placebo	114	0.0	3		2		9.1	
	CELESTIAL (24)	Cabozantinib	470	3.8	5.4	0.41 (0.34 - 0.49)	5.2	0.44 (0.36 - 0.52)	10.2	0.76 (0.63 - 0.92)
		Placebo	237	0.4	1.9		1.9		8	
	REACH-2 (25)	Ramucirumab	197	4.6	3.02	0.43 (0.31 - 0.58)	2.8	0.45 (0.34 - 0.60)	8.5	0.71 (0.53 - 0.95)
		Placebo	95	1.1	1.61		1.6		7.3	
	ADI-PEG 20 (26)	ADI-PEG 20	424	NR	NR		2.6	1.18 (0.96 - 1.43)	7.8	1.02 (0.85 - 1.23)
		Placebo	211	NR			2.6		7.4	
	ReLive (27)	Doxorubicin Transdrug	263	0.8	NR		2.3	0.95 (0.74 - 1.22)	9.1	1.00 (0.78 - 1.28)
		Placebo	134	0.7			2.3		9	

*Radiological evaluation by mRECIST
NR = Non reported
Green = Positive for superiority ($p < 0.05$)
Orange = Positive for non-inferiority (upper 95% CI < 1.08)
Red = Negative for superiority or non-inferiority

Title: Randomized trials and endpoints in advanced HCC: Role of PFS as a surrogate of survival.

Authors

Josep M Llovet[1,2,3], Robert Montal[1], Augusto Villanueva[2,4].

Affiliations

[1]Translational Research in Hepatic Oncology, Liver Unit, IDIBAPS, Hospital Clinic Barcelona, University of Barcelona, Barcelona, Catalonia, Spain.

[2]Mount Sinai Liver Cancer Program, Division of Liver Diseases, Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

[3]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain.

[4]Division of Hematology and Medical Oncology, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

Corresponding author

Josep M Llovet, Translational Research in Hepatic Oncology, IDIBAPS-Hospital Clinic, Rosselló 153, 08039, Barcelona, Catalonia, Spain; Tel. 0034-932.279.155; Email address: jmllovet@clinic.cat

Keywords: trial design; liver cancer; systemic therapies; FDA approval.

Electronic word count: 6756.

Number of figures and tables: 4 figures, 2 tables, 1 supplementary figure.

Financial support statement: JML is supported by grants from European Commission Horizon 2020 Program (HEPCAR, proposal number 667273-2), US Department of Defense (CA150272P3), National Cancer Institute (P30 CA196521), Spanish National Health Institute (SAF 2016-76390), Cancer Research UK – Asociación Española Contra el Cáncer (Accelerator Award-HUNTER-A26813), Samuel Waxman Cancer Research Foundation, Generalitat de Catalunya (AGAUR, SGR-1358). AV is supported by U.S. Department of Defense (CA150272P3) and Tisch Cancer Institute (Cancer Center Grant P30 CA196521). RM is supported by a FSEOM – Boehringer Ingelheim grant.

Conflict of interest statement: JML and AV are inventors of a patent for the immune molecular subclass in HCC. Dr Llovet is receiving research support from Bayer HealthCare Pharmaceuticals, Eisai Inc, Bristol-Myers Squibb and Ipsen, and consulting fees from Eli Lilly, Bayer HealthCare

Pharmaceuticals, Bristol-Myers Squibb, Eisai Inc, Celsion Corporation, Exelixis, Merck, Ipsen, Glycotest, Navigant, Leerink Swann LLC, Midatech Ltd, Fortress Biotech, Sprink Pharmaceuticals, Nucleix and Cat-Fite pharmaceuticals..

Summary

Hepatocellular carcinoma (HCC) is a relevant cause of cancer-related mortality worldwide. Around half of HCC patients will receive systemic therapies during their life span. The pivotal positive sorafenib trial and regulatory approval in 2007 was followed by a decade of negative studies with drugs leading to marginal anti-tumoral efficacy, toxicity, or trials with lack of enrichment strategies. This trend has changed during the period 2016-18, when several compounds such as lenvatinib (in first line) and regorafenib, cabozantinib, ramucirumab and nivolumab (in second-line) showed clinical benefit. These successes came at a cost of increasing the complexity of decision-making, and ultimately, impacting the design of future clinical trials. Nowadays, life expectancy with single active agents has surpassed the threshold of 1 year and the field is facing encouraging outcomes ~2 years with sequential strategies. Overall survival (OS) remains as the main endpoint in phase III investigations, but as in other solid tumors, there is a clear need to define surrogate endpoints that both reliably recapitulate survival benefits and can be assessed prior additional efficacious drugs are administered. A thorough analysis of 21 phase III trials published in advanced HCC demonstrated a moderate correlation between progression-free survival (PFS) or time to progression (TTP) with OS (R=0.84 and R=0.83, respectively). Nonetheless, significant differences in PFS were only followed by differences in survival in 3 out of 7 phase III studies. In these later cases, the magnitude of benefit for PFS was HR ≤ 0.6, and thus this threshold is herein proposed as a potential surrogate endpoint of OS in advanced HCC. Conversely, PFS with a HR between 0.6-0.7, despite significant, was not associated with better survival, and thus these magnitudes are considered uncertain surrogates. In the current review, we discuss the reasons for positive or negative phase III trials in advanced HCC, and the strengths and limitations of clinical surrogate endpoints (PFS, TTP and objective response rate [ORR]) to predict survival.

Introduction

The incidence of hepatocellular carcinoma (HCC) is increasing and will soon surpass one million annual cases worldwide[1]. Up to 80% of HCC patients have concomitant liver cirrhosis, mainly as result of hepatitis B and C virus infection, alcohol abuse or non-alcoholic steatohepatitis in the context of metabolic syndrome[2]. Coexistence of cancer and cirrhosis in HCC is an essential hallmark that has shaped clinical trial design in HCC, as encapsulated in the Barcelona Clinic Liver Cancer (BCLC) algorithm[3,4]. Only 40% of HCC patients are diagnosed at early stages, when potentially curative treatments (i.e. resection, liver transplantation and local ablation) are applicable[4]. As disease progresses, transarterial chemoembolization[5] (for intermediate HCC) and systemic targeted therapies[6] (for advanced HCC) have shown survival benefits. Since sorafenib positive impact in survival[7], at least 10 trials have shown negative results in front-line. In the last 2 years, however, numerous systemic agents have demonstrated clinical benefit in the context of phase III trials. The fact that six drugs are currently effective and/or approved by the Food and Drug Administration (FDA) for the management of advanced HCC poses a challenge for assessing novel strategies in this arena in terms of trial design. Overall survival (OS) is an unquestionable, unbiased primary endpoint in oncology and in all randomized studies testing systemic therapies in first and second line in advanced HCC[7-27]. However, other solid tumors have identified surrogate endpoints of survival that led to accelerated and regular approval, notably objective response rate (ORR) and progression free survival (PFS)[28,29]. These endpoints are aimed to recapitulate survival benefits with the advantage of being assessed prior additional efficacious drugs are administered. By thoroughly analyzing the past experience since sorafenib approval, we have assessed the correlation between surrogate endpoint such PFS, time to progression (TTP) and ORR with clinically meaningful improvements in OS in 21 reported phase III studies[7-27]. Based on this analysis and a conservative approach to define surrogates of OS in HCC, we propose PFS with a threshold of HR≤ 0.6 as reliable surrogate with solid positive predictive value, whereas the threshold of HR=0.6-0.7 -despite leading to positive statistical results- is defined as clinically uncertain in terms of capturing true advantages in OS. In addition, we revisit the correlation between ORR by mRECIST and OS and define again that ORR is an independent predictor of OS at early, intermediate and advanced stages, meaning that responders survive significantly longer. Nonetheless, ORR is still a suboptimal tool as surrogate due to the low sensitivity in capturing those patients that benefit from a given drug. Ultimately, we envision providing a historical perspective of HCC trial design, which are the lessons to be learned, and how to maximize clinical trials success in the near future.

Overview of phase III and practice-changing phase II trials reported during the last 10 years

Current estimates suggest that around 50% of HCC patients will receive systemic therapies at one time point or another during their lifespan [2,4,30]. Several trials have tried to show survival

benefits of systemic agents in advanced disease, a traditionally challenging setting due to the limit efficacy and high toxicity of conventional systemic chemotherapy[16–18,31]. Randomized studies also failed to prove any clinical efficacy for anti-estrogen therapies[32]. In 2008, the landmark SHARP trial assessing the multi-tyrosine kinase inhibitor sorafenib (VEGFRs, PDGFRs, RAF and KIT) was the first to significantly expand survival (hazard ratio (HR) of 0.69) with manageable adverse events[7]. Similar efficacy was demonstrated in the phase III trial testing sorafenib in Asian patients[8]. These successful results helped established contemporary concepts in trial design that have been implemented in phase III trials over the succeeding years (Table 1 and Figure 1)[33]. The main concepts implemented in these trials are: a) selection of patients with well-preserved liver function (i.e., Child-Pugh A class) to minimize the competing risk of liver failure and death as a result of the natural history of cirrhosis; b) restriction of the investigational niches to those stages with unmet medical needs such as advanced stage (BCLC C), or intermediate stage (BCLC B) progressing after TACE in case of systemic treatments or adjuvant setting after resection/local ablation; c) use of OS as the cornerstone primary endpoint to assess efficacy in advanced stages, and d) use of critical prognostic factors as tools for stratification prior randomization based upon ECOG 0 vs 1, macrovascular invasion, extrahepatic spread, and alpha-fetoprotein (AFP) levels (>400 ng/ml). Etiology is not considered a prognostic factor, but needs to be incorporated when testing sorafenib, since it has been demonstrated to be a predictor of response for this drug, at the same level than absence of extrahepatic spread and low neutrophil-to-lymphocyte ratio[34].

In this context, new treatment modalities emerged to challenge sorafenib in first-line or placebo in second-line. These include brivanib (VEGFRs and FGFRs)[10,19], sunitinib (VEGFRs, PDGFRs and KIT)[9], linifanib (VEGFRs and PDGFRs)[11], erlotinib (EGFR) in combination to sorafenib[12], everolimus (mTOR)[20], tivantinib (MET)[23], doxorubicin loaded nanoparticles[27] and ADI-PEG 20 (arginine deiminase enzyme)[26] (Table 1). All of them had disappointing results and it was not until 2016 that the RESORCE study led to the first positive phase III trial in advanced HCC for nearly a decade. Regorafenib (VEGFRs, PDGFRs, KIT and Tie2) improved OS compared to placebo from 7.8 to 10.6 months in patients who progressed and were tolerant to sorafenib[22]. Notably, OS from starting sorafenib follow by regorafenib was 26 months compared to 19 months for sorafenib followed by placebo for patients with advanced HCC [35]. Besides regorafenib, other phase III clinical trials have recently improved OS in second-line when compared to placebo: The CELESTIAL study, showing median OS of 10.2 months with cabozantinib (VEGFRs, MET and AXL) vs 8 months with placebo[24]; and the REACH-2 study, where ramucirumab (VEGFR2 monoclonal antibody) provided a median OS of 8.5 months in patients with AFP equal or higher than 400 ng/ml vs 7.3 months with placebo[21,25]. AFP is well-known for its independent prognostic capacity in HCC[36]. As such, REACH-2 becomes the first positive phase III trial in a biomarker-driven population of HCC patients. In parallel, lenvatinib (VEGFRs, FGFRs, RET, KIT and PDGFRA) has become an option in frontline after the positive result of the non-inferiority REFLECT

study[13]. In contrast, three phase III trials testing internal radiation with Y-90 for advanced HCC, either as single treatment [SARAH[14] and SIRveNIB[15]] or in combination Y-90 plus sorafenib[37] did not meet the primary endpoint of improved OS compared to sorafenib. As a result, Y-90 is discouraged for the management of advanced HCC in the recent EASL guidelines[4].

Finally, the FDA has granted accelerated approval to the immune checkpoint inhibitor nivolumab (monoclonal antibody against PD1) in second-line after a large phase II single-arm trial showing promising ORR of 14% by RECIST (responses lasting more than 12 months in 55% of cases)[38,39]. Pembrolizumab has recently shown an ORR of 17% and median OS of 12.9 months in the second-line setting[40]. The revolution of immune therapies that has changed the paradigm of treatment in oncology is now finding its way in HCC, with ongoing phase III in both first- (NCT02576509, NCT03298451, NCT03434379, NCT03713593) and second-line (NCT02702401) targeting key mediators of the anti-cancer immune response (e.g., PD1, PDL1, CTLA4, LAG3). Overall, these successful results have amplified the number of effective drugs available to clinicians for the management of advanced HCC (Figure 2). New studies will be crucial to ascertain the most efficient way to utilize these drugs and maximize clinical benefit.

Reasons for positive/negative results in phase III investigations

Negative phase III clinical trials

Until 2016, sorafenib was the only systemic agent able to significantly increase survival in patients with advanced HCC[7]. This was despite numerous attempts to improve, or parallel (i.e., non-inferiority trials), its efficacy and develop new second-line therapies in the context of phase III trials (Table 1). Until then none of the 8 randomized clinical trials (RCTs) testing systemic treatments (vs sorafenib in front line[9–12,17] or placebo in second-line[19–21]) was able to achieve positive results. Nowadays, 6 of 21 (29%) trials have been able to meet the primary endpoint and potentially change the standard of care. This success rate is lower than in other tumor types, with reported success rate of 37%[41], and resonates with the difficulties of developing effective drugs in HCC. Negative HCC trials enrolled a total of 8,604 patients and consumed a significant amount of resources. Failed drugs include linifanib[11], erlotinib[12], brivanib[10,19], sunitinib[9], doxorubicin[17,27], everolimus[20], tivantinib[23], ADI-PEG 20[26] and radioembolization with Y-90[14,15] (Table 1). These drugs have different molecular targets, mechanisms of action, and include various treatment modalities (pharmacological vs radiation-based). Thus, it is likely that multiple reasons contributed to their failure, but we will dissect three key factors, reviewed in [42]: a) limited anti-tumoral efficacy (or biological activity), b) significant toxicity, and c) lack of effective enrichment strategies for patient enrollment.

Limited anti-tumoral efficacy (or biological activity)

The first factor relates to limited anti-tumoral activity of the drug as per its main molecular targets. This implies that either they have a marginal role in HCC progression, or their selective inhibition is insufficient to induce a significant clinical benefit. For instance, evidence from murine models demonstrated how aberrant activation of MTOR signaling promotes liver cancer[43], and how its selective abrogation has anti-tumoral effects in xenografts[44]. In human HCC, MTOR pathway is deregulated in up to 45% of samples[44,45], and yet, the phase III trial testing everolimus was unable to improve survival compared to placebo in second line with a HR for OS of 1.05[20]. Data in phase II already suggested a modest median survival of 8.4 months[46] associated to marginal response rates. Also, the companion biomarker study for the phase III trial failed to find any robust predictive biomarker of response to everolimus[47]. Altogether, these data suggest that MTOR inhibition has no anti-tumoral activity in advanced HCC. Other drugs potentially falling under this category include the EGFR inhibitor erlotinib[12], and the FGFR2/VEGFR inhibitor brivanib[10,19]. Limited efficacy can also be claimed for Y-90 resin microsphere treatment as a superior alternative to sorafenib. The phase III trials reported[14,15], were negative with a HR for OS of 1.1. It is important to note that failure to demonstrate superiority does not mean similar efficacy, which requires an *ad hoc* trial design for non-inferiority or equivalence, a concept that will be further discussed in this review[48].

Drug toxicity

The second reason is significant drug toxicity, which is relevant in cirrhotic patients since liver dysfunction decreases the threshold for severe adverse effects. The best example is sunitinib[49], that despite having a molecular target profile similar to sorafenib, it was unable to improve survival when compared in frontline[9]. The trial was prematurely terminated due to futility and safety concerns affecting sunitinib. Median OS for sunitinib was 7.9 months, compared to the 10.2 with sorafenib (HR of 1.3). Treatment-related deaths occurred in 3.2% and 0.4% of patients receiving sunitinib and sorafenib, respectively. When this trial was conducted, sunitinib was already FDA approved for advanced kidney cancer and gastrointestinal stromal tumors, where toxicity was not a major clinical issue. However, in patients with underlying liver disease the toxicity of sunitinib was severe enough to obscure any beneficial anti-tumoral efficacy. Sunitinib has a higher inhibitory potency than sorafenib, particularly regarding its anti-angiogenic activity via VEGFR and PDGFR Inhibition[50]. Angiogenesis is critical during liver fibrogenesis[51], so the strong and sustained anti-angiogenic effect achieved with sunitinib seems detrimental, favoring liver failure. Previous phase II trials testing sunitinib in HCC offer additional insights into the hepatic toxicity of this drug[52–54], including up to 4/37 (11%) treatment-related deaths[54]. An adequate identification of toxicity signals at this stage could help mitigate this problem. Another example is the VEGFR/PDGFR inhibitor linifanib, tested in frontline versus sorafenib[11]. This trial was early terminated based on futility (median OS for linifanib and sorafenib were 9.1 and 9.8 months), but grade 3-4 adverse events were significantly more frequent in linifanib than in sorafenib, including hypertension (21% vs

11%) and hepatic encephalopathy (7% vs 3%). Besides the negative effect of toxicity in clinical outcomes, there is a subtler effect of non-lethal toxicity as it associates with dose reductions, which could also decrease antitumor potency.

Lack of trial enrichment strategies

A third reason for clinical trial failure is lack of effective enrichment strategies for patient enrollment based on predicted biomarkers of response. Trial enrichment in oncology is closely linked to the concept of oncogene addiction. This term describes those molecular alterations, generally DNA mutations or chromosomal aberrations, required for cancer cell proliferation and survival[55]. There are numerous examples in oncology of survival benefits after a clinical trial testing a drug only in those patients with mutations in its target[56] (e.g., ALK rearrangements in lung cancer and response to crizotinib[57]). Only 2/21 (10%) phase III trials in advanced HCC incorporated patient enrichment, likely due to: a) limited access to tumor tissue in patients already diagnosed by noninvasive criteria; and b) few druggable targets among the most common genetic alterations in HCC[58,59]. In fact, the most common mutations in HCC (*TERT* promoter, *CTNNB1*, *TP53*, *AXIN1*, *ARID1A* and *ARID1B*) are untargetable[2,60]. One of them evaluated tivantinib versus placebo in second line in patients with high expression of MET assessed with immunohistochemistry[23]. This trial was based on a post-hoc analysis of 37 patients from a previous phase 2 trial[61] and failed to meet its primary endpoint with a HR for OS of 0.97. Arguably, the signal in the phase 2 trial was weak, but most important, recent data questions the specificity of tivantinib as a MET inhibitor[62]. It was also thought that MET was a prognostic factor, but the median survival of 9.1 months for the placebo arm in MET-high patients in second-line challenges this concept[23]. The second trial tested the VEGFR2 monoclonal antibody ramucirumab versus placebo in second line (i.e. REACH-2) in patients with AFP higher than 400 ng/mL, and showed a significant improvement in OS versus placebo (HR of 0.71[25]). A difference with the tivantinib case is that the rationale for REACH-2 came from a *post hoc* analysis of the negative phase III trial in all-comers (i.e., REACH[21]) which enrolled 565 patients. This showed a robust p of interaction favoring ramucirumab in patients with high AFP of 0.02. AFP is a well-known poor prognostic marker[36,47], highly expressed in tumors with a supposed progenitor cell origin[63], but it does not provide a neat link between any specific driver oncogenic event (i.e., structural DNA alteration or signaling pathway) and ramucirumab's main molecular target. Experimental evidence identifies VEGFR2 as a marker of hepatic progenitors[64], which could hypothetically explain the efficacy of ramucirumab in tumors with high AFP.

Positive phase III clinical trials

Successful drugs in frontline include sorafenib and lenvatinib, whereas regorafenib, cabozantinib, and ramucirumab in patients with high AFP demonstrated efficacy in second line (Table 1 and Figure 2). The PD1 inhibitor nivolumab has shown promising results in phase II with an

ORR of 14% by RECIST (18% by mRECIST) and a median OS of 15.6 months [38,65], which granted its accelerated approval by the FDA. Another immune-based therapy -pembrolizumab- reported similar ORR (17%) but lower median OS (12.9 months[40]). Data from phase III trials in first-line (nivolumab vs sorafenib) and in second-line (pembrolizumab vs placebo) will be critical to fully recommend these immune-based therapies in clinical practice guidelines[4]. Since the strength of evidence so far comes from phase II data, current EASL guidelines just posed a weak recommendation for nivolumab[4].

The reasons for trial failure provide the best clues for the qualities of a drug to be successful in HCC, which essentially are: a) adequate clinical trial design with an emphasis on selection criteria and robust endpoints; b) a fine balance between drug efficacy and toxicity; and c) a proper interpretation of efficacy and toxicity signals in phase II trials. Sorafenib epitomizes these qualities, and to certain extent, the design principles implemented in the pivotal SHARP trial[7] were adopted as best-practices for design in subsequent studies[33]. The target population must include patients with well-preserved liver function (i.e., Child-Pugh A with compensated liver disease) to avoid competing risks from deaths due to progression of the liver disease, and to minimize drug toxicity. Also, patients need to be fit enough to tolerate the drug and with a life expectancy of at least 3 months, which can be reasonably guaranteed by enrolling patients with ECOG performance status test (PST) of 0-1. It is paramount to enroll patients at the same clinical stage as per the BCLC classification[4,66]. The SHARP trial was instrumental to eradicate the misleading concept of 'unresectable' HCC when conducting HCC trials. This concept included a heterogeneous population of patients at intermediate (BCLC-B) and advanced (BCLC-B) stages[67], which imposed significant bias when interpreting trial results. In addition to the same clinical stage, patients need to be adequately stratified for known HCC prognostic factors and geographic region.

Regarding patient's selection, the success of the REACH-2 trial underscores the importance of properly interpreting *post hoc* analysis. The pooled analysis of REACH and REACH-2 assessing ramucirumab in those patients with AFP > 400 ng/mL further confirms a significant and clinically meaningful benefit of ramucirumab vs placebo in second-line (median survival 8.1 vs 5 months; HR=0.694)[25]. The rationale to enrich trials based on predicted oncogene addiction is two-fold: first, to maximize anti-tumoral response by perturbing the cancer drivers active in a given patient, and second, to spare unnecessary toxicity in those patients without the oncogene addiction. HCC has few druggable targets among the most frequent driver mutations, but a recent proof-of-concept trial reinforces the validity of this approach to explore treatment response[68]. A screening of 1,318 HCC patients allowed enrolling 54 cases with RAS mutations (4.4%) detected using circulating tumor DNA (ctDNA) in a phase II trial testing the combination of sorafenib with the MEK inhibitor refametinib[68]. Mutation analysis of ctDNA is feasible in HCC[69] and facilitates screening of large populations. Other potential druggable oncogenic alterations in HCC include high-level DNA amplifications of *FGF19*[70,71] or *VEGFA*[70,72]. Phase II clinical trials are currently exploring

selective inhibition of these candidate oncogene addition loops[73]. There is also increasing interest in developing biomarkers to identify the 20% of patients who respond to immune-based therapies, who show outstanding OS. The use of PD-L1 staining seems irrelevant in HCC[38], and other potential biomarkers such as tumor mutational burden (TMB)[74] or the HCC immune class[75] are under investigation. To facilitate the implementation of biomarker-based clinical trials in HCC, it is essential to enforce mandatory tissue collection in all clinical trials testing new compounds[4]. In this regard, the impact of intratumor heterogeneity in single-biopsy predictions is still debated despite recent studies have shown that driver gene mutations are common between different regions of the tumor[76,77].

Traditionally, new therapies were compared with standard of care or placebo to demonstrate greater efficacy of the new drug. Despite this is the recommended trial design practice in HCC[4,33], some studies after SHARP used non-inferiority designs to challenge sorafenib in first-line. The hypothesis in non-inferiority trials is that the new compound is not substantially worse than the current standard, as opposed to equivalence trials, which are designed to demonstrate that the experimental treatment is neither worse not better than the standard therapy[78]. Non-inferiority trials are required to claim similar efficacy as opposed to assuming it from a negative superiority trial, as previously explained for the Y-90 trials. The non-inferiority trial scenario in HCC is extensively described elsewhere[42], and caveats include the need for larger sample sizes and a very small window of opportunity, as defined by the tight non-inferiority margins. For instance, the BRISK-FL trial was designed to demonstrate non-inferiority of brivanib compared to sorafenib in first line[10]. The trial assumptions set the upper limit of the 95% confidence interval of HR for OS to 1.08. To call non-inferiority, the HR could cross 1, but the upper boundary needed to fall between 1 and 1.08. This threshold is very stringent and can be interpreted as the requirement to demonstrate a robust non-significant trend towards superiority for the new drug. The value proposed by FDA has been calculated based upon capturing at least >60% of the survival benefit obtained with sorafenib[13]. The BRISK-FL trial did not meet this endpoint since the HR confidence interval limits for OS were 0.94 and 1.23. The concept of non-inferiority trials introduces other considerations in treatment recommendations such as toxicity or cost, which will surely contribute to frame the landscape of systemic therapies in HCC.

Hard and surrogate endpoints: Implications in clinical trial design

The overreaching goal of oncological treatments is to allow patients to live longer and better lives than they would do without treatment[79]. Thus, clinical research needs to unequivocally demonstrate statistically and clinically meaningful improvements of the experimental arm over the standard of care. Three types of endpoints have been defined: 1. Hard endpoints, such as overall survival and cancer-specific survival; 2. Surrogate endpoints such as PFS, TTP and ORR, and 3. Patient reported endpoints, such as quality of life (QOL).

Overall survival

This hard endpoint quantifies the time between random trial allocation and death, whatever the cause. Since is not subject to investigator bias, OS has been traditionally recommended by international HCC guidelines as the primary endpoint for randomized phase III trials testing new therapies[80]. In fact, all regular FDA drug approvals in advanced HCC were based upon improvements in OS [7,22]. Cancer-specific survival, where only deaths due to cancer are considered and noncancer-related deaths are censored, is more difficult to assess in conventional trial settings. Deaths due to competing risks, such as liver failure, require a subjective interpretation by the investigator, and thus are more prone to bias[33].

What is the magnitude of benefit to define it as clinically meaningful? This is a controversial topic, highly dependent on the expected outcome for the target population, with confronted opinions between patients, providers, regulatory agencies and health insurers [81,82]. In HCC, there is no consensus on what absolute survival benefit (or the magnitude of benefit in OS as per HR) to be defined as clinically relevant. Reported thresholds of OS with HR<0.8 are also sound for capturing the benefit of patients in advanced HCC trials[83].

Survival has some limitations as a sole endpoint in cancer research. First, it might require a long follow-up time to capture enough events due to significant improvement in median OS in the experimental arm[35]. This negatively impacts feasibility and delays access to patients to highly effective drugs. Second, it can be affected by sequential therapies received after tumor progression (post-progression survival), such as for instance regorafenib after a first-line therapy. This might involve one third of patients in recent phase III trials[13]. In this context, validation of surrogate endpoints of OS is paramount to facilitate trial execution and favor a quick deployment of effective drugs in routine clinical conditions.

Surrogate endpoints: PFS, TTP and ORR

Ideally, significant improvement in OS is preferred, but many drugs have been approved based on their ability to improve other less robust endpoints, termed surrogates (i.e., TTP, PFS and ORR). These are outcomes not inherently meaningful from the clinical standpoint, but thought to accurately predict hard outcomes such as OS[84]. The development of surrogate endpoints became a necessity in clinical trials in cardiology, where the long time to accumulate enough events for a hard endpoint made most studies unfeasible. Use of surrogate endpoints is becoming a need in oncology where effective post-progression therapies are available.

Accelerated approval based upon surrogate endpoints is becoming the most relevant path for cancer drug regulatory approval in the US. Between 2009 and 2014; the FDA approved 83 drugs in oncology, 66% of them on the basis of surrogate endpoints[85]. The FDA's accelerated approval program was introduced in 1992 as a social compromise during the worse years of the HIV

epidemic to expedite access to agents for life-threating conditions based on surrogate endpoints. The program included a "safety net" that required the manufacturer to conduct post-marketing studies and confirm the efficacy of the drug using hard endpoints[86]. A recent analysis of approved drugs during the period 1992-2017 led to the following conclusions[28]: a) Accelerated approval was granted for 93 indications, ORR being the most common surrogate endpoint used (87% of cases), b) Among drugs approved through this path, 55% were ultimately confirmed for regular approval, 5% of indications were withdrawn (e.g., bevacizumab in metastatic breast cancer[87]), whereas in others the process has not been concluded.

Despite the increasing importance of surrogate endpoints in oncology, they have two main limitations. First, since they usually rely on the radiological definition of tumor progression or response, they are vulnerable to interpretation bias. This can be minimized by using central radiology reviews and a designated adjudicator of response. Second, and more important, in order to be reliable, they require validation as credible predictors of OS[84]. Validation of surrogate endpoints can be conducted at the individual- or trial-levels[88]. While validation of individual-level surrogacy requires individual patient data from at least one clinical trial, trial-level surrogacy uses assembled data from multiple trials. The Institute for Quality and Efficiency in Health Care has proposed a set of criteria to quantify the association between a surrogate and hard endpoint, which includes low (R<0.7), moderate (R>0.7 to R<0.85) and high correlation (R>0.85)[89]. R refers to the weighted Pearson coefficient between HR of OS and HR of the surrogate endpoint. Alternative methods to study this correlation have been reported[90]. A systematic review and meta-analysis of trial-level surrogate endpoints (PFS, TTP and ORR) for OS in oncology including 36 articles and 352 clinical trials found low, moderate and high correlation with OS in 52%, 25% and 23% of surrogate endpoints, respectively[84].

In order to explore the concept of surrogate endpoints recapitulating OS in advanced HCC we have identified 21 RCTs assessing systemic therapies with or without loco-regional therapies in advanced HCC (12 in first-line and 9 in second-line) (Table 1) published between 2008 and 2018 through a MEDLINE search via PubMed using the keywords "advanced hepatocellular carcinoma". Results were limited to "clinical trial, phase III". Trials recently presented at international meetings (2016-2018) were also included despite the full manuscript is not yet available. For each trial, data on sample size, radiological response, TTP and OS were collected. TTP and OS were determined in terms of hazard ratio (HR) using published data (values less than 1 denotes a favourable result in the experimental group). In addition, ORR was established with odds ratio calculated from the published radiological response (values greater than 1 denotes a favourable result in the experimental group). For the purpose of the trial-level analysis, we first assessed the overall correlation between PFS and OS (R=0.84; R^2=0.71) (Figure 3A), and then the correlation of TTP and OS (R=0.83; R^2=0.69) (Figure 3B). Afterwards, we established a conservative threshold of positive predictive value for PFS since this is the most documented surrogate time-to-event

endpoint in oncology, and the one showing a higher correlation with OS[84]. Finally, we explored the correlation between ORR assessed as per mRECIST and survival in early, intermediate and advanced HCC.

Progression-free survival

PFS is a composite endpoint of two variables: death and evidence of radiological progression, usually defined by standard criteria as RECIST[91] or mRECIST[92]. International guidelines initially discouraged this endpoint in HCC due to the competing risk of dying due to progressed liver dysfunction despite a relevant anti-tumoral benefit[33]. However, this limitation has been mitigated since most trials in HCC have adopted restrictive inclusion criteria in terms of liver function (i.e., Child-Pugh A without decompensation). In this scenario, the likelihood of death as a result of liver decompensation (i.e., gastrointestinal bleeding, encephalopathy or ascites and spontaneous peritonitis) is 5% at one year[93]. When we evaluate the association between PFS and OS in HCC phase III trials, we observe a moderate Pearson correlation (R) of 0.84 (Figure 3A). This figure falls in the upper boundary of a moderate correlation (R between 0.7 and 0.85). When specifically analyzing the positive predictive value of theoretical thresholds of PFS correlating with OS, only 3/7 PFS reported a HR ≤0.6 that was significantly associated with a positive survival clinical benefit (in all cases with a HR for OS < 0.8). Conversely, those four studies reporting a positive PFS with a HR between 0.6-0.8 were associated with no significant survival benefits (HR for OS between 0.87 and 1.05) (Figure 3A). In our study, according to the linear regression equation obtained [log HR_{OS} = 0.072 + 0.487 x log HR_{PFS}], a threshold of PFS HR=0.6 (representing a 40% risk reduction) will decrease ~17% the risk of OS (OS HR=0.83) (see Figure 3A). In summary, a moderate correlation has been established between PFS and OS in 21 RCTs in advanced HCC. A value of HR ≤0.6 is proposed as surrogate threshold effect [94], and is likely to predict a clinically meaningful improvement in OS. Is worth to mention though that this rule is supported mainly by positive trials in the second-line setting comparing active drugs vs placebo. We assume that such association is retained in front-line comparing two active drugs, but recommendation in that setting should be tempered due to the lack of confirmatory data.

Time to progression

This endpoint quantifies the time between trial allocation and radiological progression, usually defined by standard criteria as RECIST[91] or mRECIST[92]. Deaths are censored as non-radiological progressions at the time of death or at an earlier visit, with a cause-specific hazard, representing informative censoring. Symmetric repeated radiological measurements every 6-8 weeks are required to avoid missing moderate differences between treatment groups[33]. This recommendation was not followed in the SIRveNIB[15], SARAH[14] and ADI-PEG 20[26] trials.

To delineate the adequacy of TTP as a surrogate of OS in HCC we conducted also a trial-level meta-analysis to evaluate the correlation between TTP and OS in 21 RCTs (Figure 3B). The

Pearson correlation (R) was 0.83, which indicates a moderate association according to the IQWIG guidelines[84]. In 10 phase III trials there was a significant difference in TTP in favor of the investigational arm. However, these positive results in TTP were not followed by superiority in OS in 5 (50%) trials. Brivanib[19] and ramucirumab[21] in second-line showed efficacy as per TTP (HR=0.56 and 0.59, respectively), while not significantly improving the hard endpoint of OS (HR=0.89 and 0.87, respectively). Lenvatinib[13] in first-line versus sorafenib also showed significant differences in TTP favoring lenvatinib (HR 0.63), without showing superiority for OS (HR=0.92). This trial was positive since it was designed for non-inferiority (upper 95% CI lower than 1.08). Finally, linifanib[11] and hepatic infusion arterial chemotherapy (HAIC)[18] in first-line failed to show any benefit in terms of OS (HR=1.05 and 1.01, respectively) even though there was a clear benefit when measuring TTP (HR=0.76 and 0.65, respectively). These results do not support the initial recommendation after the SHARP trial of using TTP as the optimal surrogate endpoint[33] in phase 2 trials, and reinforce the need for accurate evaluation of surrogacy in clinical trials. Based on the linear regression model obtained [log HR_{OS} = 0.083 + 0.491 x log HR_{TTP}], we can extrapolate that a therapy producing a 40% risk reduction in TTP will yield an estimated ~16% risk reduction in OS (HR=0.84) (see Figure 3B). Moreover, in order to directly compare the performance of PFS and TTP, we analyzed the correlation between both surrogate endpoints, obtaining a Pearson correlation (R) of 0.99 (Figure 4). Thus, in the modern era of HCC trial design, with minimal cirrhosis-related deaths (due to the inclusion of Child-Pugh A), there is a strong correlation between both endpoints. In fact, when we inferred the non-reported PFS HR of SHARP and AP trials according to the linear equation obtained comparing both surrogate endpoints [log HR_{PFS} = 0.014 + 0.927 x log HR_{TTP}], the HRs values are close to 0.60, just at the previously proposed minimum threshold.

There are two other considerations regarding this endpoint. First, not all types of tumor progression may have the same clinical meaning. Recent data also suggest that TTP may capture heterogeneous features, with essentially two types of progression at advanced stages[95,96]. In particular, survival after progression is significantly worse for patients who develop a new extrahepatic lesion and/or vascular invasion (median OS = 7.1 months) compared to those who progress due to the growth of existing intrahepatic/extrahepatic lesions or the development of a new intrahepatic lesion (median OS = 14.9 months). Second, factors including evaluation bias, trial attrition or informative censoring may weaken the association between the TTP and OS [29]. Finally, prolonged exposure to a given therapy might lead to a phenotypic change in tumors, thus, offsetting any initial advantage from the treatment captured by the surrogate endpoint[97].

Objective response rate

Tumor response in oncology trials is typically measured using the Response Evaluation Criteria in Solid Tumors (RECIST)[91]. These criteria standardize methods for converting radiological observations into a quantitative and statistically tractable framework to define tumor

response (i.e., a 30% decrease in the diameter of target lesion). ORR is the percentage of patients who achieve an objective tumor response. Disease control rate (DCR) is the combination of ORR and stable disease, but it has two disadvantages that limit its adoption for regulatory approval: a) the definition of duration of stable disease varies between studies; and b) stable disease can reflect inherent characteristics of the tumor rather than treatment efficacy.

The RECIST criteria were originally developed to evaluate cytotoxic agents. The generalization of targeted therapies has challenged this simplistic approach that relies on tumor shrinkage to indicate clinical efficacy. Sorafenib was associated with only 2-3% of ORR, despite providing clear survival benefits[7,8]. Given the poor correlation between tumor response assessed with conventional tools and OS, a group of experts convened by the American Association for the Study of Liver Diseases (AASLD) proposed specific amendments to standard RECIST[33]. Further description of response and progression resulted in the criteria named modified RECIST (mRECIST), which ultimately incorporates the concept of viable tumor defined as the portions of tumor showing arterial enhancement[92]. The mRECIST criteria in HCC have improved the sensitivity to quantify tumor response with targeted therapies: ORR of 9-17% with sorafenib[10,13,18], 10-12% with brivanib[10,19], 11% with regorafenib[22] and 24% with lenvatinib[13]. Retrospective studies have consistently demonstrated that patients who achieved an objective response on sorafenib had a longer survival than non-responders[98–100]. Recently, data from double-blind randomized trials assessing brivanib and nintedanib further validated this association [101,102]. Thus, the association between tumor response and improved OS in HCC patients at advanced stages complement what was already knew in patients at early and intermediate stages treated with loco-regional therapies[103–109](Table 2).

When we evaluate the trial-level correlation between ORR and OS (Supplementary Figure 1), the R weighted Pearson coefficient obtained is 0.54. This is significantly lower than the correlation obtained with PFS/TTP and OS as depicted in Figure 3. There are two reasons for this: one is inherent to the use of odds ratio instead of hazard ratio to compare differences in ORR. The accuracy of odds ratio decreases for low values of ORRs. The second reason is that only a small proportion of patients within these trials achieved ORR (~10-20%), which is, in fact, the event that correlates with better survival[101]. A direct comparison between RECIST and mRECIST for OS surrogacy through an independent meta-analysis of trials using either criteria would be ideal to define the role of ORR to predict OS in advanced HCC. This will also help determine the best tool to evaluate tumor response to systemic therapies. However, since only 5 RCTs reported response data using mRECIST, we did not sub-analyze this endpoint according to the tool used to evaluate response.

Some other questions remain unanswered. As observed in other solid tumors treated with efficacious targeted therapies[110–112], the reported rates of responders are still suboptimal to estimate the maximum number of patients who would benefit from the treatment. In addition, the

duration of response might be more clinically relevant than the extent of tumor reduction. Finally, the strategy to evaluate response might require a thoughtful revision when assessing immunotherapies. As shown in melanoma patients treated with checkpoint inhibitors, standard RECIST may not provide a reliable assessment of antitumor efficacy[113]. In fact, response to immunotherapy may take longer compared to other agents and can even falsely mirror criteria for progression (i.e., pseudo-progression)[114]. Immune-related response criteria have been developed[115,116], including the concept of "confirmation of progression" by a second scan obtained at least 4 weeks after progressive disease has been registered.

Despite all the challenges that evaluation of tumor response face in oncology, and particularly in HCC, the importance of ORR as a surrogate endpoint is recognized by regulatory agencies and frequently used for accelerated drug approval. This was the case of nivolumab, approved in second-line based on an ORR of 18% by mRECIST and 14% by RECIST [38,39]. Remarkably, objective response to nivolumab has been associated with prolonged OS[65]. Overall, the fact that a high ORR in phase II trials was considered a robust criterion for drug approval[117], and further success in phase III trials[118], indicates that ORR should be considered as a primary endpoint for single-arm phase II studies. Related to this, early clinical trials are showing promising results with combinations of checkpoint inhibitors and targeted therapies, as measured by ORR. Lenvatinib plus pembrolizumab[119] and atezolizumab plus bevacizumab[120] achieved an ORR of 46% by mRECIST and 50% by RECIST in advanced HCC, respectively. As a result, the later combination was granted breakthrough therapy designation by the FDA[121]. Of note, most of the drugs approved under the accelerated program reported ORR exceeding 30%[122].

Patient-reported endpoints: Quality of life

Health-related QOL measures the effect of the disease on an individual's physical, psychological and social functioning and well-being[123]. Regulatory agencies recognize symptomatic improvement as a direct clinical benefit to patients and an important consideration in drug approval[124]. However, unlike OS, the interpretation of QOL is subjective. In HCC, two tools have been proposed to measure QOL: the European Organization for Research and Treatment of Cancer Quality of Live Questionnaire (EORTC QLQ-HCC18)[125] and the Functional Assessment of Cancer Therapy-Hepatobiliary (FACT-Hep) questionnaire[126]. They can be used to evaluate time to symptomatic progression (i.e. time between trial allocation and the occurrence of disease related symptoms according to preestablished scores). QOL was measured in the SHARP trial[7] according to the FHSI-8 questionnaire[127], a reduced version of FACT-Hep, with results that collided with OS. Conversely, the SARAH trial[14] reported better global health status with Y-90 when compared to sorafenib based on QLQ-HCC18, which was inconsistent with the primary endpoint of OS. Evaluation of QOL is contingent on when it was assessed during disease progression. Also, significant changes in QOL have been observed across different cultures[128].

Defining and evaluating reliable QOL assessment tools has been established as one of the unmet needs in HCC research by international guidelines[4]. In summary, health-related QOL measures are not ready to support, as single tools, regulatory approval for drugs in HCC.

Conclusions

The current period of drug development in HCC is providing major advancements in the management of this devastating disease. Six drugs have currently shown activity as systemic therapies, which represents an unprecedented revolution for the last 50 years. Novel drugs or combinations strategies are emerging in the field, and thus new tools will be required for the proper assessment of clinical benefits. OS is still the most robust endpoint but the increasing number of treatments available in advanced HCC preludes the use of surrogate endpoints, less vulnerable to subsequent treatments after progression. In this scenario, PFS has shown moderate correlation with OS (R=0.84), and a threshold of HR≤0.6 defines a conservative approach of surrogate endpoint able to capture survival differences in a superiority trial with a high positive predictive value. Two recent studies have been released supporting our threshold of HR ≤ 0.6 for PFS. The first one, an individual-patient data meta-analysis of two RCTs (REACH[21] and REACH-2[25]), showing a significant OS HR with a PFS HR of 0.57[129]. The second one, a phase 3 RCT comparing sorafenib with or without cTACE in advanced HCC, a negative study for its primary endpoint (OS) with a PFS HR of 0.73[130]. Thus, PFS-HR ≤0.6 could be considered a candidate endpoint in phase II and phase III RCTs when subsequent therapies are expected to impact overall outcome. ORR by sensitive criteria (mRECIST) may be useful particularly in single arm phase II trials with proof of concept drugs or in combination studies targeting accelerated approval with a threshold >30%. Finally, the current development of RCTs assessing immune therapies or drug combinations in HCC will certainly evolve the paradigm of drug development and trial design. Particularly of interest will be whether the statements proposed in the present review are confirmed in trials designed with composite primary endpoints, such as OS-PFS for lenvatinib+pembrolizumab vs lenvatinib (NCT03713593) or OS-ORR for atezolizumab+bevacizumab vs sorafenib (NCT03434379).

Key points

- In the last two years four systemic agents (i.e., regorafenib, lenvatinib, cabozantinib and ramucirumab) have shown clinical benefit in the setting of phase III trials and one (i.e., nivolumab) has been granted accelerated approval based on a phase II trial, expanding, thus, the pipeline of effective drugs available in advanced HCC to providers.
- The improvement in the number of effective agents comes at a cost of increased complexity of clinical decision-making, and thus, in the design of future clinical trials.

- OS is still the most robust endpoint in advanced HCC but the increasing number of treatments after progression underscore the need for surrogate endpoints.
- PFS has a moderate correlation at trial level with OS (R=0.84). A conservative minimum surrogate threshold effect of HR ≤0.6 is highly predictive of a significant improvement in OS, whereas HR ranging from 0.6 – 0.7 are uncertain surrogates.
- ORR by sensitive criteria in single arm phase II trials could be a useful tool to prioritize treatments for testing in phase III trials.

References

[1] Torre L, Bray F, Siegel R, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA Cancer J Clin 2015;65:87–108.

[2] Llovet JM, Zucman-Rossi J, Pikarsky E, Sangro B, Schwartz M, Sherman M, et al. Hepatocellular carcinoma. Nat Rev Dis Prim 2016;2:16018.

[3] Llovet JM, Brú C, Bruix J. Prognosis of hepatocellular carcinoma: the BCLC staging classification. Semin Liver Dis 1999;19:329–38.

[4] Galle PR, Forner A, Llovet JM, Mazzaferro V, Piscaglia F, Raoul J, et al. EASL Clinical Practice Guidelines: Management of hepatocellular carcinoma. J Hepatol 2018;69:182–236.

[5] Llovet JM, Bruix J. Systematic review of randomized trials for unresectable hepatocellular carcinoma: Chemoembolization improves survival. Hepatology 2003;37:429–42.

[6] Llovet JM, Montal R, Sia D, Finn RS. Molecular therapies and precision medicine for hepatocellular carcinoma. Nat Rev Clin Oncol 2018;In press.

[7] Llovet J, Ricci S, Mazzaferro V, Hilgard P, Gane E, Blanc J, et al. Sorafenib in advanced hepatocellular carcinoma. N Engl J Med 2008;359:378–90.

[8] Cheng A-L, Kang Y-K, Chen Z, Tsao C-J, Qin S, Kim JS, et al. Efficacy and safety of sorafenib in patients in the Asia-Pacific region with advanced hepatocellular carcinoma: a phase III randomised, double-blind, placebo-controlled trial. Lancet Oncol 2009;10:25–34.

[9] Cheng A-L, Kang Y-K, Lin D-Y, Park J-W, Kudo M, Qin S, et al. Sunitinib versus sorafenib in advanced hepatocellular cancer: results of a randomized phase III trial. J Clin Oncol 2013;31:4067–75.

[10] Johnson PJ, Qin S, Park J-W, Poon RTP, Raoul J-L, Philip PA, et al. Brivanib versus sorafenib as first-line therapy in patients with unresectable, advanced hepatocellular carcinoma: results from the randomized phase III BRISK-FL study. J Clin Oncol 2013;31:3517–24.

[11] Cainap C, Qin S, Huang W-T, Chung IJ, Pan H, Cheng Y, et al. Linifanib versus Sorafenib in patients with advanced hepatocellular carcinoma: results of a randomized phase III trial. J

Clin Oncol 2015;33:172–9.

[12] Zhu a. X, Rosmorduc O, Evans TRJ, Ross PJ, Santoro A, Carrilho FJ, et al. SEARCH: A Phase III, Randomized, Double-Blind, Placebo-Controlled Trial of Sorafenib Plus Erlotinib in Patients With Advanced Hepatocellular Carcinoma. J Clin Oncol 2014;33:559–66.

[13] Kudo M, Finn R, Qin S, Han SS, Ikeda K, Piscaglia F, et al. A Randomised Phase 3 Trial of Lenvatinib vs. Sorafenib in Firstline Treatment of Patients With Unresectable Hepatocellular Carcinoma. Lancet 2018;391:1163–73.

[14] Vilgrain V, Pereira H, Assenat E, Guiu B, Ilonca AD, Pageaux GP, et al. Efficacy and safety of selective internal radiotherapy with yttrium-90 resin microspheres compared with sorafenib in locally advanced and inoperable hepatocellular carcinoma (SARAH): an open-label randomised controlled phase 3 trial. Lancet Oncol 2017;18:1624–36.

[15] Chow PKH, Gandhi M, Tan S-B, Khin MW, Khasbazar A, Ong J, et al. SIRveNIB: Selective Internal Radiation Therapy Versus Sorafenib in Asia-Pacific Patients With Hepatocellular Carcinoma. J Clin Oncol 2018;36:1913–21.

[16] Qin S, Bai Y, Lim HY, Thongprasert S, Chao Y, Fan J, et al. Randomized, multicenter, open-label study of oxaliplatin plus fluorouracil/leucovorin versus doxorubicin as palliative chemotherapy in patients with advanced hepatocellular carcinoma from Asia. J Clin Oncol 2013;31:3501–8.

[17] Abou-Alfa GK, Niedzwieski D, Knox JJ, Kaubisch A, Posey J, Tan BR, et al. Phase III randomized study of sorafenib plus doxorubicin versus sorafenib in patients with advanced hepatocellular carcinoma (HCC): CALGB 80802 (Alliance). J Clin Oncol 2016;34:4_suppl.

[18] Kudo M, Ueshima K, Yokosuka O, Ogasawara S, Obi S, Izumi N, et al. Sorafenib plus low-dose cisplatin and fluorouracil hepatic arterial infusion chemotherapy versus sorafenib alone in patients with advanced hepatocellular carcinoma (SILIUS): a randomised, open label, phase 3 trial. Lancet Gastroenterol Hepatol 2018;3:424–32.

[19] Llovet JM, Decaens T, Raoul JL, Boucher E, Kudo M, Chang C, et al. Brivanib in patients with advanced hepatocellular carcinoma who were intolerant to sorafenib or for whom sorafenib failed: results from the randomized phase III BRISK-PS study. J Clin Oncol 2013;31:3509–16.

[20] Zhu AX, Kudo M, Assenat E, Cattan S, Kang Y-K, Lim HY, et al. Effect of everolimus on survival in advanced hepatocellular carcinoma after failure of sorafenib: the EVOLVE-1 randomized clinical trial. JAMA 2014;312:57–67.

[21] Zhu AX, Park JO, Ryoo B-Y, Yen C-J, Poon R, Pastorelli D, et al. Ramucirumab versus placebo as second-line treatment in patients with advanced hepatocellular carcinoma following first-line therapy with sorafenib (REACH): a randomised, double-blind, multicentre, phase 3 trial. Lancet Oncol 2015:859–70.

[22] Bruix J, Qin S, Merle P, Granito A, Huang Y-H, Bodoky G, et al. Regorafenib for patients with

hepatocellular carcinoma who progressed on sorafenib treatment (RESORCE): a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet 2017;389:56–66.

[23] Rimassa L, Assenat E, Peck-Radosavljevic M, Zagonel V, Pracht M, Rota Caremolli E. Second-line tivantinib (ARQ 197) vs placebo in patients (Pts) with MET-high hepatocellular carcinoma (HCC): Results of the METIV-HCC phase III trial. J Clin Oncol 2017;35:15_supp.

[24] Abou-Alfa GK, Meyer T, Cheng A-L, El-Khoueiry AB, Rimassa L, Ryoo B-Y, et al. Cabozantinib in Patients with Advanced and Progressing Hepatocellular Carcinoma. N Engl J Med 2018;379:54–63.

[25] Zhu AX, Kang Y-K, Yen C-J, Finn RS, Galle PR, Llovet JM, et al. REACH-2: A randomized, double-blind, placebo-controlled phase 3 study of ramucirumab versus placebo as second-line treatment in patients with advanced hepatocellular carcinoma (HCC) and elevated baseline alpha-fetoprotein (AFP) following first-line sorafe. J Clin Oncol 2018;36 Suppl.

[26] Abou-Alfa GK, Qin S, Lu S-N, Yen C-J, Feng Y-H, Lim Y, et al. Phase III randomized study of second line ADI-peg 20 (A) plus best supportive care versus placebo (P) plus best supportive care in patients (pts) with advanced hepatocellular carcinoma (HCC). Ann Oncol 2018;29:1402–8.

[27] Merle P, Bodoky G, López-López C, Saad AS, Casadei Gardini A, Borbath I, et al. Safety and Efficacy Results from the Phase 3 Relive Study of Doxorubicin-Loaded Nanoparticles versus Best Standard of Care in Patients with Advanced Hepatocellular Carcinoma after Failure or Intolerance to Previous Treatment Including Sorafenib. Int Liver Cancer Assoc 11th Annu Conf 2017.

[28] Beaver JA, Howie LJ, Pelosof L, Kim T, Liu J, Goldberg KB, et al. A 25-Year Experience of US Food and Drug Administration Accelerated Approval of Malignant Hematology and Oncology Drugs and Biologics. JAMA Oncol 2018;4:849–56.

[29] Kemp R, Prasad V. Surrogate endpoints in oncology: When are they acceptable for regulatory and clinical decisions, and are they currently overused? BMC Med 2017;15:1–7.

[30] Bruix J SMAA for the S of LD. Management of hepatocellular carcinoma: an update. Hepatology 2011;53:1020–2.

[31] Yeo W, Mok TS, Zee B, Leung TWT, Lai PBS, Lau WY, et al. A randomized phase III study of doxorubicin versus cisplatin/interferon alpha-2b/doxorubicin/fluorouracil (PIAF) combination chemotherapy for unresectable hepatocellular carcinoma. J Natl Cancer Inst 2005;97:1532–8.

[32] Chow P, Tai B-C, Tan C-K, Machin D, Win KM, Johnson PJ, et al. High-dose tamoxifen in the treatment of inoperable hepatocellular carcinoma: A multicenter randomized controlled trial. Hepatology 2002;36:1221–6.

[33] Llovet JM, Di Bisceglie AM, Bruix J, Kramer BS, Lencioni R, Zhu AX, et al. Design and endpoints of clinical trials in hepatocellular carcinoma. J Natl Cancer Inst 2008;100:698–711.

[34] Bruix J, Cheng A-L, Meinhardt G, Nakajima K, De Sanctis Y, Llovet J. Prognostic Factors and Predictors of Sorafenib Benefit in Patients With Hepatocellular Carcinoma: Analysis of Two Phase 3 Studies. J Hepatol 2017;67:999–1008.

[35] Finn RS, Merle P, Granito A, Huang Y-H, Bodoky G, Pracht M, et al. Outcomes with sorafenib followed by regorafenib or placebo for HCC: additional analyses from the phase 3 RESORCE trial. J Hepatol 2018;In press.

[36] Llovet JM, Pena CE a., Lathia CD, Shan M, Meinhardt G, Bruix J. Plasma Biomarkers as Predictors of Outcome in Patients with Advanced Hepatocellular Carcinoma. Clin Cancer Res 2012;18:2290–300.

[37] Ricke J, Sangro B, Amthauer H, Bargellini I, Bartenstein P, De Toni E, et al. The impact of combining Selective Internal Radiation Therapy (SIRT) with sorafenib on overall survival in patients with advanced hepatocellular carcinoma: The SORAMIC trial palliative cohort. J Hepatol 2018;68 | S65–S.

[38] El-Khoueiry AB, Sangro B, Yau T, Crocenzi TS, Kudo M, Hsu C, et al. Nivolumab in patients with advanced hepatocellular carcinoma (CheckMate 040): An open-label, non-comparative, phase 1/2 dose escalation and expansion trial. Lancet 2017;389:2492–502.

[39] FDA grants accelerated approval to nivolumab for HCC previously treated with sorafenib n.d. https://www.fda.gov/Drugs/InformationOnDrugs/ApprovedDrugs/ucm577166.htm.

[40] Zhu AX, Finn RS, Edeline J, Cattan S, Ogasawara S, Palmer D, et al. Pembrolizumab in patients with advanced hepatocellular carcinoma previously treated with sorafenib (KEYNOTE-224): a non-randomised, open-label phase 2 trial. Lancet Oncol 2018;19.

[41] Gan HK, You B, Pond GR, Chen EX. Assumptions of expected benefits in randomized phase III trials evaluating systemic treatments for cancer. J Natl Cancer Inst 2012;104:590–8.

[42] Llovet JM, Hernandez-Gea V. Hepatocellular carcinoma: reasons for phase III failure and novel perspectives on trial design. Clin Cancer Res 2014;20:2072–9.

[43] Guri Y, Colombi M, Dazert E, Hindupur SK, Roszik J, Moes S, et al. mTORC2 Promotes Tumorigenesis via Lipid Synthesis. Cancer Cell 2017;32:807–823.e12.

[44] Villanueva A, Chiang DY, Newell P, Peix J, Thung S, Alsinet C, et al. Pivotal Role of mTOR Signaling in Hepatocellular Carcinoma. Gastroenterology 2008;135:1972–83.

[45] Sahin F, Kannangai R, Adegbola O, Wang J, Su G, Torbenson M. mTOR and P70 S6 kinase expression in primary liver neoplasms. Clin Cancer Res 2004;10:8421–5.

[46] Zhu AX, Abrams TA, Miksad R, Blaszkowsky LS, Meyerhardt JA, Zheng H, et al. Phase 1/2 study of everolimus in advanced hepatocellular carcinoma. Cancer 2011;117:5094–102.

[47] Zhu AX, Chen D, He W, Kanai M, Voi M, Chen LT, et al. Integrative biomarker analyses indicate etiological variations in hepatocellular carcinoma. J Hepatol 2016;65:296–304.

[48] Llovet JM, Finn RS. Negative phase 3 study of 90Y microspheres versus sorafenib in HCC. Lancet Oncol 2018;19:e69.

[49] Faivre S, Delbaldo C, Vera K, Robert C, Lozahic S, Lassau N, et al. Safety, pharmacokinetic, and antitumor activity of SU11248, a novel oral multitarget tyrosine kinase inhibitor, in patients with cancer. J Clin Oncol 2006;24:25–35.

[50] Stein MN, Flaherty KT. Sorafenib and sunitinib in renal cell carcinoma. Clin Cancer Res 2007;13:3765–70.

[51] Poisson J, Lemoinne S, Boulanger C, Durand F, Moreau R, Valla D, et al. Liver sinusoidal endothelial cells: Physiology and role in liver diseases. J Hepatol 2017;66:212–27.

[52] Koeberle D, Montemurro M, Samaras P, Majno P, Simcock M, Limacher A, et al. Continuous Sunitinib Treatment in Patients with Advanced Hepatocellular Carcinoma: A Swiss Group for Clinical Cancer Research (SAKK) and Swiss Association for the Study of the Liver (SASL) Multicenter Phase II Trial (SAKK 77/06). Oncologist 2010;15:285–92.

[53] Zhu AX, Sahani D V., Duda DG, Di Tomaso E, Ancukiewicz M, Catalano OA, et al. Efficacy, safety, and potential biomarkers of sunitinib monotherapy in advanced hepatocellular carcinoma: A phase II study. J Clin Oncol 2009;27:3027–35.

[54] Faivre S, Raymond E, Boucher E, Douillard J, Lim HY, Kim JS, et al. Safety and efficacy of sunitinib in patients with advanced hepatocellular carcinoma: an open-label, multicentre, phase II study. Lancet Oncol 2009;10:794–800.

[55] Luo J, Solimini NL, Elledge SJ. Principles of Cancer Therapy: Oncogene and Non-oncogene Addiction. Cell 2009;136:823–37.

[56] Villanueva A, Llovet JM. Targeted therapies for hepatocellular carcinoma. Gastroenterology 2011;140:1410–26.

[57] Shaw AT, Kim D-W, Nakagawa K, Seto T, Crinó L, Ahn M-J, et al. Crizotinib versus Chemotherapy in Advanced *ALK* -Positive Lung Cancer. N Engl J Med 2013;368:2385–94.

[58] Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. Nat Med 2017;23:703–13.

[59] Schulze K, Nault J-C, Villanueva A. Genetic profiling of hepatocellular carcinoma using next-generation sequencing. J Hepatol 2016;65:1031–42.

[60] Zucman-Rossi J, Villanueva A, Nault JC, Llovet JM. Genetic Landscape and Biomarkers of Hepatocellular Carcinoma. Gastroenterology 2015;149:1226–39.

[61] Santoro A, Rimassa L, Borbath I, Daniele B, Salvagni S, Van Laethem JL, et al. Tivantinib for second-line treatment of advanced hepatocellular carcinoma: A randomised, placebo-controlled phase 2 study. Lancet Oncol 2013;14:55–63.

[62] Rebouissou S, La Bella T, Rekik S, Imbeaud S, Calatayud AL, Rohr-Udilova N, et al. Proliferation markers are associated with MET expression in hepatocellular carcinoma and predict tivantinib sensitivity in vitro. Clin Cancer Res 2017;23:4364–75.

[63] Hoshida Y, Nijman SMB, Kobayashi M, Chan JA, Brunet J-P, Chiang DY, et al. Integrative

transcriptome analysis reveals common molecular subclasses of human hepatocellular carcinoma. Cancer Res 2009;69:7385–92.

[64] Goldman O, Han S, Sourrisseau M, Dziedzic N, Hamou W, Corneo B, et al. KDR identifies a conserved human and murine hepatic progenitor and instructs early liver development. Cell Stem Cell 2013;12:748–60.

[65] El-Khoueiry AB, Melero I, Yau TC, Crocenzi TS, Kudo M, Hsu C, et al. Impact of antitumor activity on survival outcomes, and nonconventional benefit, with nivolumab (NIVO) in patients with advanced hepatocellular carcinoma (aHCC): Subanalyses of CheckMate-040. J Clin Oncol 2018;36, suppl.

[66] Forner A, Reig M, Bruix J. Hepatocellular carcinoma. Lancet 2018;391:1301–14.

[67] Llovet JM, Bustamante J, Castells a, Vilana R, Ayuso MDC, Sala M, et al. Natural history of untreated nonsurgical hepatocellular carcinoma: rationale for the design and evaluation of therapeutic trials. Hepatology 1999;29:62–7.

[68] Lim HY, Merle P, Weiss KH, Yau T, Ross PJ, Mazzaferro V, et al. Phase II Studies with Refametinib or Refametinib plus Sorafenib in Patients with RAS-mutated Hepatocellular Carcinoma. Clin Cancer Res 2018;June 27.

[69] Labgaa I, Villacorta-Martin C, D'Avola D, Craig AJ, von Felden J, Martins-Filho SN, et al. A pilot study of ultra-deep targeted sequencing of plasma DNA identifies driver mutations in hepatocellular carcinoma. Oncogene 2018:1–13.

[70] Chiang DY, Villanueva A, Hoshida Y, Peix J, Newell P, Minguez B, et al. Focal gains of VEGFA and molecular classification of hepatocellular carcinoma. Cancer Res 2008;68:6779–88.

[71] Sawey ET, Chanrion M, Cai C, Wu G, Zhang J, Zender L, et al. Identification of a therapeutic strategy targeting amplified FGF19 in liver cancer by Oncogenomic screening. Cancer Cell 2011;19:347–58.

[72] Horwitz E, Stein I, Andreozzi M, Nemeth J, Shoham A, Pappo O, et al. Human and mouse VEGFA-amplified hepatocellular carcinomas are highly sensitive to Sorafenib treatment. Cancer Discov 2014;4:730–43.

[73] Kang Y-K, Macarulla T, Yau T, Sarker D, Choo SP, Meyer T, et al. Clinical Activity of Blu-554, a Potent, Highly-Selective FGFR4 Inhibitor in Advanced Hepatocellular Carcinoma (HCC) with FGFR4 Pathway Activation. Int Liver Cancer Assoc 11th Annu Conf 2017.

[74] Hellmann MD, Ciuleanu T-E, Pluzanski A, Lee JS, Otterson GA, Audigier-Valette C, et al. Nivolumab plus Ipilimumab in Lung Cancer with a High Tumor Mutational Burden. N Engl J Med 2018;378:2093–104.

[75] Sia D, Jiao Y, Martinez-Quetglas I, Kuchuk O, Villacorta-Martin C, Castro De Moura M. Identification of an Immune-specific Class of Hepatocellular Carcinoma, Based on Molecular Features. Gastroenterology 2017;153:812–26.

[76] Torrecilla S, Sia D, Harrington AN, Zhang Z, Cabellos L, Cornella H, et al. Trunk mutational events present minimal intra- and inter-tumoral heterogeneity in hepatocellular carcinoma. J Hepatol 2017;67:1222–31.

[77] Reiter JG, Makohon-Moore AP, Gerold JM, Heyde A, Attiyeh MA, Kohutek ZA, et al. Minimal functional driver gene heterogeneity among untreated metastases. Science 2018;361:1033–7.

[78] Kaji AH, Lewis RJ. Noninferiority Trials: Is a New Treatment Almost as Effective as Another? Jama 2015;313:2371–2.

[79] Wilson MK, Karakasis K, Oza AM. Outcomes and endpoints in trials of cancer treatment: the past, present, and future. Lancet Oncol 2015;16:e32–42.

[80] Llovet JM, Di Bisceglie AM, Bruix J, Kramer BS, Lencioni R, Zhu AX, et al. Design and endpoints of clinical trials in hepatocellular carcinoma. J Natl Cancer Inst 2008;100:698–711.

[81] Booth CM, Tannock I. Reflections on medical oncology: 25 years of clinical trials--where have we come and where are we going? J Clin Oncol 2008;26:6–8.

[82] Cherny NI, Dafni U, Bogaerts J, Latino NJ, Pentheroudakis G, Douillard JY, et al. ESMO-Magnitude of Clinical Benefit Scale version 1.1. Ann Oncol 2017;28:2340–66.

[83] Ellis LM, Bernstein DS, Voest EE, Berlin JD, Sargent D, Cortazar P, et al. American society of clinical oncology perspective: Raising the bar for clinical trials by defining clinically meaningful outcomes. J Clin Oncol 2014;32:1277–80.

[84] Prasad V, Kim C, Burotto M, Vandross A. The strength of association between surrogate end points and survival in oncology: A systematic review of trial-level meta-analyses. JAMA Intern Med 2015;175:1389–98.

[85] Kim C, Prasad V. Strength of Validation for Surrogate End Points Used in the US Food and Drug Administration's Approval of Oncology Drugs. Mayo Clin Proc 2016;91:713–25.

[86] Gyawali B, Kesselheim AS. Reinforcing the social compromise of accelerated approval. Nat Rev Clin Oncol 2018;July 3.

[87] Carpenter D, Kesselheim AS, Joffe S. Reputation and Precedent in the Bevacizumab Decision. N Engl J Med 2011;365:e3.

[88] Zhao F. Surrogate End Points and Their Validation in Oncology Clinical Trials. J Clin Oncol 2016;34:1436–7.

[89] Institute for Quality and Efficiency in Health Care [Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen] (IQWiG). Validity of surrogate endpoints in oncology: executive summary. n.d. http: //www.iqwig.de/download/A10-05_Executive _Summary_v1-1_Surrogate_endpoints_in_oncology.pdf.

[90] Ciani O, Buyse M, Garside R, Peters J, Saad ED, Stein K, et al. Meta-analyses of randomized controlled trials show suboptimal validity of surrogate outcomes for overall survival in advanced colorectal cancer. J Clin Epidemiol 2015;68:833–42.

[91] Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). Eur J Cancer 2009;45:228–47.

[92] Lencioni R, Llovet J. Modified RECIST (mRECIST) Assessment for Hepatocellular Carcinoma. Semin Liver Dis 2010;30:052–60.

[93] D'Amico G, Garcia-Tsao G, Pagliaro L. Natural history and prognostic indicators of survival in cirrhosis: A systematic review of 118 studies. J Hepatol 2006;44:217–31.

[94] Burzykowski T, Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. Pharm Stat 2006;5:173–86.

[95] Reig M, Rimola J, Torres F, Darnell A, Rodriguez-Lope C, Forner A, et al. Postprogression survival of patients with advanced hepatocellular carcinoma: rationale for second-line trial design. Hepatology 2013;58:2023–31.

[96] Iavarone M, Cabibbo G, Biolato M, Della Corte C, Maida M, Barbara M, et al. Predictors of survival in patients with advanced hepatocellular carcinoma who permanently discontinued sorafenib. Hepatology 2015;62:784–91.

[97] Booth CM, Eisenhauer EA, Group NCT. Progression-free survival: meanigful or simply measurable? J Clin Oncol 2012;30:1030–3.

[98] Edeline J, Boucher E, Rolland Y, Vauléon E, Pracht M, Perrin C, et al. Comparison of tumor response by Response Evaluation Criteria in Solid Tumors (RECIST) and modified RECIST in patients treated with sorafenib for hepatocellular carcinoma. Cancer 2012;118:147–56.

[99] Ronot M, Bouattour M, Wassermann J, Bruno O, Dreyer C, Larroque B, et al. Alternative Response Criteria (Choi, European association for the study of the liver, and modified Response Evaluation Criteria in Solid Tumors [RECIST]) Versus RECIST 1.1 in patients with advanced hepatocellular carcinoma treated with sorafenib. Oncologist 2014;19:394–402.

[100] Takada J, Hidaka H, Nakazawa T, Kondo M, Numata K, Tanaka K, et al. Modified response evaluation criteria in solid tumors is superior to response evaluation criteria in solid tumors for assessment of responses to sorafenib in patients with advanced hepatocellular carcinoma. BMC Res Notes 2015;8:609.

[101] Lencioni R, Montal R, Torres F, Park J-W, Decaens T, Raoul J-L, et al. Objective response by mRECIST as a predictor and potential surrogate end-point of overall survival in advanced HCC. J Hepatol 2017;66:1166–72.

[102] Meyer T, Palmer DH, Cheng A-L, Hocke J, Loembé A-B, Yen C-J. mRECIST to predict survival in advanced hepatocellular carcinoma: Analysis of two randomised phase II trials comparing nintedanib vs sorafenib. Liver Int 2017:1–9.

[103] Cabibbo G, Maida M, Genco C, Alessi N, Peralta M, Butera G, et al. Survival of Patients with Hepatocellular Carcinoma (HCC) Treated by Percutaneous Radio-Frequency Ablation (RFA) Is Affected by Complete Radiological Response. PLoS One 2013;8.

[104] Vincenzi B, Di Maio M, Silletta M, D'Onofrio L, Spoto C, Piccirillo MC, et al. Prognostic Relevance of Objective Response According to EASL Criteria and mRECIST Criteria in Hepatocellular Carcinoma Patients Treated with Loco-Regional Therapies: A Literature-Based Meta-Analysis. PLoS One 2015;10:e0133488.

[105] Sala M, Llovet JM, Vilana R, Bianchi L, Solé M, Ayuso C, et al. Initial response to percutaneous ablation predicts survival in patients with hepatocellular carcinoma. Hepatology 2004;40:1352–60.

[106] Gillmore R, Stuart S, Kirkwood A, Hameeduddin A, Woodward N, Burroughs AK, et al. EASL and mRECIST responses are independent prognostic factors for survival in hepatocellular cancer patients treated with transarterial embolization. J Hepatol 2011;55:1309–16.

[107] Kim BK, Kim KA, Park JY, Ahn SH, Chon CY, Han KH, et al. Prospective comparison of prognostic values of modified Response Evaluation Criteria in Solid Tumours with European Association for the Study of the Liver criteria in hepatocellular carcinoma following chemoembolisation. Eur J Cancer 2013;49:826–34.

[108] Jung ES, Kim JHJS, Yoon EL, Lee HJ, Lee SJ, Suh SJ, et al. Comparison of the methods for tumor response assessment in patients with hepatocellular carcinoma undergoing transarterial chemoembolization. J Hepatol 2013;58:1181–7.

[109] Prajapati HJ, Spivey JR, Hanish SI, El-rayes BF, Kauh JS, Chen Z, et al. mRECIST and EASL responses at early time point by contrast-enhanced dynamic MRI predict survival in patients with unresectable hepatocellular carcinoma (HCC) treated by doxorubicin drug-eluting beads transarterial chemoembolization (DEB TACE). Ann Oncol 2013;24:965–73.

[110] Hudes G, Carducci M, Tomczak P, Dutcher J, Figlin R, Kapoor A, et al. Temsirolimus, interferon alpha or both for advanced renal cell carcinoma. N Engl J Med 2007;356:2271–81.

[111] Karapetis CS, Khambata-Ford S, Jonker DJ, O'Callaghan CJ, Tu D, Tebbutt NC, et al. K-ras mutations and benefit from cetuximab in advanced colorectal cancer. N Engl J Med 2008;359:1757–65.

[112] Hurwitz H, Fehrenbacher L, Novotny W, Cartwright T, Hainsworth J, Heim W, et al. Bevacizumab plus Irinotecan, Fluorouracil, and Leucovorin for Metastatic Colorectal Cancer. N Engl J Med 2004;350:2335–42.

[113] Hodi FS, Hwu WJ, Kefford R, Weber JS, Daud A, Hamid O, et al. Evaluation of immune-related response criteria and RECIST v1.1 in patients with advanced melanoma treated with Pembrolizumab. J Clin Oncol 2016;34:1510–7.

[114] Wolchok JD, Hoos A, O'Day S, Weber JS, Hamid O, Lebbé C, et al. Guidelines for the evaluation of immune therapy activity in solid tumors: Immune-related response criteria. Clin Cancer Res 2009;15:7412–20.

[115] Seymour L, Bogaerts J, Perrone A, Ford R, Schwartz LH, Mandrekar S, et al. iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics. Lancet Oncol

2017;18:e143–52.

[116] Stephen Hodi F, Ballinger M, Lyons B, Soria JC, Nishino M, Tabernero J, et al. Immune-modified response evaluation criteria in solid tumors (imrecist): Refining guidelines to assess the clinical benefit of cancer immunotherapy. J Clin Oncol 2018;36:850–8.

[117] Tsimberidou AM, Braiteh F, Stewart DJ, Kurzrock R. Ultimate fate of oncology drugs approved by the US food and drug administration without a randomized trial. J Clin Oncol 2009;27:6243–50.

[118] El-Maraghi RH, Eisenhauer EA. Review of Phase II Trial Designs Used in Studies of Molecular Targeted Agents: Outcomes and Predictors of Success in Phase III. J Clin Oncol 2008;26:1346–54.

[119] Ikeda M, Sung MW, Kudo M, Kobayashi M, Baron AD, Finn RS, et al. A phase 1b trial of lenvatinib (LEN) plus pembrolizumab (PEM) in patients (pts) with unresectable hepatocellular carcinoma (uHCC). J Clin Oncol 2018;36, suppl.

[120] Stein S, Pishvaian MJ, Lee MS, Lee K-H, Hernandez S, Kwan A, et al. Safety and clinical activity of 1L atezolizumab + bevacizumab in a phase Ib study in hepatocellular carcinoma (HCC). J Clin Oncol 2018;36, suppl.

[121] FDA Grants Breakthrough Therapy Designation for Genentech's TECENTRIQ in Combination With Avastin as First-Line Treatment for Advanced or Metastatic Hepatocellular Carcinoma (HCC) 2018. https://www.gene.com/media/press-releases/14736/2018-07-17/fda-grants-breakthrough-therapy-designat?utm_source=F&utm_medium=P&utm_term=15538&utm_content=TecentriqHCCBTD&utm_campaign=TecentriqHCCBTD.

[122] Oxnard GR, Wilcox KH, Gonen M, Polotsky M, Hirsch BR, Schwartz LH. Response rate as a regulatory end point in single-arm studies of advanced solid tumors. JAMA Oncol 2016;2:772–9.

[123] Revicki DA, Osoba D, Fairclough D, Barofsky I, Berzon R, Leidy NK, et al. Recommendations on health-related quality of life research to support labeling and promotional claims in the United States. Qual Life Res 2000;9:887–900.

[124] Wilson MK, Collyar D, Chingos DT, Friedlander M, Ho TW, Karakasis K, et al. Outcomes and endpoints in cancer trials: Bridging the divide. Lancet Oncol 2015;16:e43–52.

[125] Chie W-C, Blazeby JM, Hsiao C-F, Chiu H-C, Poon RT, Mikoshiba N, et al. International cross-cultural field validation of an European Organization for Research and Treatment of Cancer questionnaire module for patients with primary liver cancer, the European Organization for Research and Treatment of Cancer quality-of-life ques. Hepatology 2012;55:1122–9.

[126] Heffernan N, Cella D, Webster K, Odom L, Martone M, Passik S, et al. Measuring health-related quality of life in patients with hepatobiliary cancers: The functional assessment of

Cancer Therapy-Hepatobiliary Questionnaire. J Clin Oncol 2002;20:2229–39.

[127] Yount S, Cella D, Webster K, Heffernan N, Chang CH, Odom L, et al. Assessment of patient-reported clinical outcome in pancreatic and other hepatobiliary cancers: The FACT Hepatobiliary Symptom Index. J Pain Symptom Manage 2002;24:32–44.

[128] Chie W-C, Blazeby JM, Hsiao C-F, Chiu H-C, Poon RT, Mikoshiba N, et al. Differences in health-related quality of life between European and Asian patients with hepatocellular carcinoma. Asia Pac J Clin Oncol 2017:e304–11.

[129] Llovet JM, Kudo M, Finn R, Galle PR, Blanc J, Okusaka T, et al. Ramucirumab As Second-Line Treatment in Patients with Hepatocellular Carcinoma (HCC) and Elevated Alpha-Fetoprotein (AFP) Following Sorafenib: Pooled Results from Two Global Phase 3 Studies (REACH-2 and REACH). Hepatology 2018;68, suppl.

[130] Park J-W, Kim YJ, Kim DY, Bae SH, Paik SW, Lee Y-J. Sorafenib with or without concurrent transarterial chemoembolization in patients with advanced hepatocellular carcinoma: a phase III STAH trial. J Hepatol 2018;In Press.

Figure legends

Figure 1: Median overall survival of treatment modalities assessed in phase III trials for advanced hepatocellular carcinoma.

Treatments with more than one dot represent all the results obtained from different clinical trials testing the same compound. Trials are colored based on whether the final result was positive for superiority (green), negative (red) or positive for non-inferiority (orange) for the primary endpoint (OS). Placebo appears in blue. Relevant inclusion/exclusion criteria that may impact on median OS are: no portal vein invasion [13], no pulmonary metastases [37], sorafenib tolerant [22], MET high [23] and AFP>400ng/ml [25].

Figure 2: Treatment strategy for advanced hepatocellular carcinoma. Adapted from Llovet et al. Nat Rev Clin Oncol 2018[6]**.**

Drugs in green have positive results from phase III trials with a superiority design (sorafenib in the first-line setting and regorafenib, cabozantinib and ramucirumab in the second-line setting). Drugs in orange have positive results from phase III trials with a non-inferiority design (lenvatinib in the first-line setting). Drugs in red have received accelerated approval from the FDA on the basis of promising efficacy results in phase II trials (nivolumab in the second-line setting). Key details of the patient populations are provided. AFP: Alpha-fetoprotein; BCLC: Barcelona Clinic Liver Cancer (classification); ECOG PS: Eastern Cooperative Oncology Group performance status; EHS: extrahepatic spread; HCV: hepatitis C virus; HR: hazard ratio; mRECIST: modified Response Evaluation Criteria In Solid Tumors; ORR: objective response rate; OS: overall survival.

Figure 3: Correlation between surrogate endpoints (PFS[A] and TTP[B]) and hard endpoint (OS).

Trial-level correlation between endpoints using criteria from the Institute for Quality and Efficiency in Health Care (IQWIG). R and R2 refers to the weighted Pearson coefficient between the HR of OS and the HR of the surrogate endpoint. IQWIG categorizes the strength of the correlation based on the value of R as low (R<0.7), moderate (R>0.7 to R<0.85) and high (R>0.85)[89]. Each dot represents one of the phase III clinical trials conducted on advanced HCC. Size of the dot is proportional to the total number of patients enrolled in the trial. Trials are colored based on whether the final result was positive, negative or non-inferior for the primary endpoint (OS). X and Y axis depict the value of the HR for the surrogate (TTP or PFS) and the hard endpoint (OS), respectively. Gray shaded areas represent the upper and lower limits of the 95% confidence intervals for the regression.

Figure 4: Correlation between surrogate endpoints PFS and TTP.

Trial-level correlation between endpoints. R and R2 refers to the weighted Pearson coefficient between the HR of PFS and the HR of TTP. Each dot represents one of the phase III clinical trials conducted on advanced HCC. Size of the dot is proportional to the total number of patients enrolled in the trial. Trials are colored based on whether the final result was positive, negative or non-inferior for the primary endpoint (OS). X and Y axis depict the value of the HR for the surrogate TTP and PFS, respectively. Gray shaded areas represent the upper and lower limits of the 95% confidence intervals for the regression.

Supplementary Figure 1: Correlation between objective response and OS.

Trial-level correlation between endpoints. R and R2 refers to the weighted Pearson coefficient between the HR of OS and the odds ratio of objective response. Each dot represents one of the phase III clinical trials conducted on advanced HCC. Size of the dot is proportional to the total number of patients enrolled in the trial. Trials are colored based on whether the final result was positive, negative or non-inferior for the primary endpoint (OS). X and Y axis depict the value of the odds ratio for objective response and the HR for OS, respectively. Gray shaded areas represent the upper and lower limits of the 95% confidence intervals for the regression.

Tables

Table 1: Phase III trials in advanced hepatocellular carcinoma conducted in the last decade.

Table 2: Studies analyzing associations between radiological response and survival in hepatocellular carcinoma.

Title: Randomized trials and endpoints in advanced HCC: Role of PFS as a surrogate of survival.

Authors

Josep M Llovet[1,2,3], Robert Montal[1], Augusto Villanueva[2,4].

Affiliations

[1]Translational Research in Hepatic Oncology, Liver Unit, IDIBAPS, Hospital Clinic Barcelona, University of Barcelona, Barcelona, Catalonia, Spain.

[2]Mount Sinai Liver Cancer Program, Division of Liver Diseases, Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

[3]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain.

[4]Division of Hematology and Medical Oncology, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

Corresponding author

Josep M Llovet, Translational Research in Hepatic Oncology, IDIBAPS-Hospital Clinic, Rosselló 153, 08039, Barcelona, Catalonia, Spain; Tel. 0034-932.279.155; Email address: jmllovet@clinic.cat

Keywords: trial design; liver cancer; systemic therapies; FDA approval.

Electronic word count: 6756.

Number of figures and tables: 4 figures, 2 tables, 1 supplementary figure.

Financial support statement: JML is supported by grants from European Commission Horizon 2020 Program (HEPCAR, proposal number 667273-2), US Department of Defense (CA150272P3), National Cancer Institute (P30 CA196521), Spanish National Health Institute (SAF 2016-76390), Cancer Research UK – Asociación Española Contra el Cáncer (Accelerator Award-HUNTER-A26813), Samuel Waxman Cancer Research Foundation, Generalitat de Catalunya (AGAUR, SGR-1358). AV is supported by U.S. Department of Defense (CA150272P3) and Tisch Cancer Institute (Cancer Center Grant P30 CA196521). RM is supported by a FSEOM – Boehringer Ingelheim grant.

Conflict of interest statement: JML and AV are inventors of a patent for the immune molecular subclass in HCC. Dr Llovet is receiving research support from Bayer HealthCare Pharmaceuticals, Eisai Inc, Bristol-Myers Squibb and Ipsen, and consulting fees from Eli Lilly, Bayer HealthCare

Pharmaceuticals, Bristol-Myers Squibb, Eisai Inc, Celsion Corporation, Exelixis, Merck, Ipsen, Glycotest, Navigant, Leerink Swann LLC, Midatech Ltd, Fortress Biotech, Sprink Pharmaceuticals, Nucleix and Cat-Fite pharmaceuticals..

Summary

Hepatocellular carcinoma (HCC) is a relevant cause of cancer-related mortality worldwide. Around half of HCC patients will receive systemic therapies during their life span. The pivotal positive sorafenib trial and regulatory approval in 2007 was followed by a decade of negative studies with drugs leading to marginal anti-tumoral efficacy, toxicity, or trials with lack of enrichment strategies. This trend has changed during the period 2016-18, when several compounds such as lenvatinib (in first line) and regorafenib, cabozantinib, ramucirumab and nivolumab (in second-line) showed clinical benefit. These successes came at a cost of increasing the complexity of decision-making, and ultimately, impacting the design of future clinical trials. Nowadays, life expectancy with single active agents has surpassed the threshold of 1 year and the field is facing encouraging outcomes ~2 years with sequential strategies. Overall survival (OS) remains as the main endpoint in phase III investigations, but as in other solid tumors, there is a clear need to define surrogate endpoints that both reliably recapitulate survival benefits and can be assessed prior additional efficacious drugs are administered. A thorough analysis of 21 phase III trials published in advanced HCC demonstrated a moderate correlation between progression-free survival (PFS) or time to progression (TTP) with OS (R=0.84 and R=0.83, respectively). Nonetheless, significant differences in PFS were only followed by differences in survival in 3 out of 7 phase III studies. In these later cases, the magnitude of benefit for PFS was HR ≤ 0.6, and thus this threshold is herein proposed as a potential surrogate endpoint of OS in advanced HCC. Conversely, PFS with a HR between 0.6-0.7, despite significant, was not associated with better survival, and thus these magnitudes are considered uncertain surrogates. In the current review, we discuss the reasons for positive or negative phase III trials in advanced HCC, and the strengths and limitations of clinical surrogate endpoints (PFS, TTP and objective response rate [ORR]) to predict survival.

Introduction

The incidence of hepatocellular carcinoma (HCC) is increasing and will soon surpass one million annual cases worldwide[1]. Up to 80% of HCC patients have concomitant liver cirrhosis, mainly as result of hepatitis B and C virus infection, alcohol abuse or non-alcoholic steatohepatitis in the context of metabolic syndrome[2]. Coexistence of cancer and cirrhosis in HCC is an essential hallmark that has shaped clinical trial design in HCC, as encapsulated in the Barcelona Clinic Liver Cancer (BCLC) algorithm[3,4]. Only 40% of HCC patients are diagnosed at early stages, when potentially curative treatments (i.e. resection, liver transplantation and local ablation) are applicable[4]. As disease progresses, transarterial chemoembolization[5] (for intermediate HCC) and systemic targeted therapies[6] (for advanced HCC) have shown survival benefits. Since sorafenib positive impact in survival[7], at least 10 trials have shown negative results in front-line. In the last 2 years, however, numerous systemic agents have demonstrated clinical benefit in the context of phase III trials. The fact that six drugs are currently effective and/or approved by the Food and Drug Administration (FDA) for the management of advanced HCC poses a challenge for assessing novel strategies in this arena in terms of trial design. Overall survival (OS) is an unquestionable, unbiased primary endpoint in oncology and in all randomized studies testing systemic therapies in first and second line in advanced HCC[7-27]. However, other solid tumors have identified surrogate endpoints of survival that led to accelerated and regular approval, notably objective response rate (ORR) and progression free survival (PFS)[28,29]. These endpoints are aimed to recapitulate survival benefits with the advantage of being assessed prior additional efficacious drugs are administered. By thoroughly analyzing the past experience since sorafenib approval, we have assessed the correlation between surrogate endpoint such PFS, time to progression (TTP) and ORR with clinically meaningful improvements in OS in 21 reported phase III studies[7-27]. Based on this analysis and a conservative approach to define surrogates of OS in HCC, we propose PFS with a threshold of HR≤ 0.6 as reliable surrogate with solid positive predictive value, whereas the threshold of HR=0.6-0.7 -despite leading to positive statistical results- is defined as clinically uncertain in terms of capturing true advantages in OS. In addition, we revisit the correlation between ORR by mRECIST and OS and define again that ORR is an independent predictor of OS at early, intermediate and advanced stages, meaning that responders survive significantly longer. Nonetheless, ORR is still a suboptimal tool as surrogate due to the low sensitivity in capturing those patients that benefit from a given drug. Ultimately, we envision providing a historical perspective of HCC trial design, which are the lessons to be learned, and how to maximize clinical trials success in the near future.

Overview of phase III and practice-changing phase II trials reported during the last 10 years

Current estimates suggest that around 50% of HCC patients will receive systemic therapies at one time point or another during their lifespan [2,4,30]. Several trials have tried to show survival

benefits of systemic agents in advanced disease, a traditionally challenging setting due to the limit efficacy and high toxicity of conventional systemic chemotherapy[16–18,31]. Randomized studies also failed to prove any clinical efficacy for anti-estrogen therapies[32]. In 2008, the landmark SHARP trial assessing the multi-tyrosine kinase inhibitor sorafenib (VEGFRs, PDGFRs, RAF and KIT) was the first to significantly expand survival (hazard ratio (HR) of 0.69) with manageable adverse events[7]. Similar efficacy was demonstrated in the phase III trial testing sorafenib in Asian patients[8]. These successful results helped established contemporary concepts in trial design that have been implemented in phase III trials over the succeeding years (Table 1 and Figure 1)[33]. The main concepts implemented in these trials are: a) selection of patients with well-preserved liver function (i.e., Child-Pugh A class) to minimize the competing risk of liver failure and death as a result of the natural history of cirrhosis; b) restriction of the investigational niches to those stages with unmet medical needs such as advanced stage (BCLC C), or intermediate stage (BCLC B) progressing after TACE in case of systemic treatments or adjuvant setting after resection/local ablation; c) use of OS as the cornerstone primary endpoint to assess efficacy in advanced stages, and d) use of critical prognostic factors as tools for stratification prior randomization based upon ECOG 0 vs 1, macrovascular invasion, extrahepatic spread, and alpha-fetoprotein (AFP) levels (>400 ng/ml). Etiology is not considered a prognostic factor, but needs to be incorporated when testing sorafenib, since it has been demonstrated to be a predictor of response for this drug, at the same level than absence of extrahepatic spread and low neutrophil-to-lymphocyte ratio[34].

In this context, new treatment modalities emerged to challenge sorafenib in first-line or placebo in second-line. These include brivanib (VEGFRs and FGFRs)[10,19], sunitinib (VEGFRs, PDGFRs and KIT)[9], linifanib (VEGFRs and PDGFRs)[11], erlotinib (EGFR) in combination to sorafenib[12], everolimus (mTOR)[20], tivantinib (MET)[23], doxorubicin loaded nanoparticles[27] and ADI-PEG 20 (arginine deiminase enzyme)[26] (Table 1). All of them had disappointing results and it was not until 2016 that the RESORCE study led to the first positive phase III trial in advanced HCC for nearly a decade. Regorafenib (VEGFRs, PDGFRs, KIT and Tie2) improved OS compared to placebo from 7.8 to 10.6 months in patients who progressed and were tolerant to sorafenib[22]. Notably, OS from starting sorafenib follow by regorafenib was 26 months compared to 19 months for sorafenib followed by placebo for patients with advanced HCC [35]. Besides regorafenib, other phase III clinical trials have recently improved OS in second-line when compared to placebo: The CELESTIAL study, showing median OS of 10.2 months with cabozantinib (VEGFRs, MET and AXL) vs 8 months with placebo[24]; and the REACH-2 study, where ramucirumab (VEGFR2 monoclonal antibody) provided a median OS of 8.5 months in patients with AFP equal or higher than 400 ng/ml vs 7.3 months with placebo[21,25]. AFP is well-known for its independent prognostic capacity in HCC[36]. As such, REACH-2 becomes the first positive phase III trial in a biomarker-driven population of HCC patients. In parallel, lenvatinib (VEGFRs, FGFRs, RET, KIT and PDGFRA) has become an option in frontline after the positive result of the non-inferiority REFLECT

study[13]. In contrast, three phase III trials testing internal radiation with Y-90 for advanced HCC, either as single treatment [SARAH[14] and SIRveNIB[15]] or in combination Y-90 plus sorafenib[37] did not meet the primary endpoint of improved OS compared to sorafenib. As a result, Y-90 is discouraged for the management of advanced HCC in the recent EASL guidelines[4].

Finally, the FDA has granted accelerated approval to the immune checkpoint inhibitor nivolumab (monoclonal antibody against PD1) in second-line after a large phase II single-arm trial showing promising ORR of 14% by RECIST (responses lasting more than 12 months in 55% of cases)[38,39]. Pembrolizumab has recently shown an ORR of 17% and median OS of 12.9 months in the second-line setting[40]. The revolution of immune therapies that has changed the paradigm of treatment in oncology is now finding its way in HCC, with ongoing phase III in both first- (NCT02576509, NCT03298451, NCT03434379, NCT03713593) and second-line (NCT02702401) targeting key mediators of the anti-cancer immune response (e.g., PD1, PDL1, CTLA4, LAG3). Overall, these successful results have amplified the number of effective drugs available to clinicians for the management of advanced HCC (Figure 2). New studies will be crucial to ascertain the most efficient way to utilize these drugs and maximize clinical benefit.

Reasons for positive/negative results in phase III investigations

Negative phase III clinical trials

Until 2016, sorafenib was the only systemic agent able to significantly increase survival in patients with advanced HCC[7]. This was despite numerous attempts to improve, or parallel (i.e., non-inferiority trials), its efficacy and develop new second-line therapies in the context of phase III trials (Table 1). Until then none of the 8 randomized clinical trials (RCTs) testing systemic treatments (vs sorafenib in front line[9–12,17] or placebo in second-line[19–21]) was able to achieve positive results. Nowadays, 6 of 21 (29%) trials have been able to meet the primary endpoint and potentially change the standard of care. This success rate is lower than in other tumor types, with reported success rate of 37%[41], and resonates with the difficulties of developing effective drugs in HCC. Negative HCC trials enrolled a total of 8,604 patients and consumed a significant amount of resources. Failed drugs include linifanib[11], erlotinib[12], brivanib[10,19], sunitinib[9], doxorubicin[17,27], everolimus[20], tivantinib[23], ADI-PEG 20[26] and radioembolization with Y-90[14,15] (Table 1). These drugs have different molecular targets, mechanisms of action, and include various treatment modalities (pharmacological vs radiation-based). Thus, it is likely that multiple reasons contributed to their failure, but we will dissect three key factors, reviewed in [42]: a) limited anti-tumoral efficacy (or biological activity), b) significant toxicity, and c) lack of effective enrichment strategies for patient enrollment.

Limited anti-tumoral efficacy (or biological activity)

The first factor relates to limited anti-tumoral activity of the drug as per its main molecular targets. This implies that either they have a marginal role in HCC progression, or their selective inhibition is insufficient to induce a significant clinical benefit. For instance, evidence from murine models demonstrated how aberrant activation of MTOR signaling promotes liver cancer[43], and how its selective abrogation has anti-tumoral effects in xenografts[44]. In human HCC, MTOR pathway is deregulated in up to 45% of samples[44,45], and yet, the phase III trial testing everolimus was unable to improve survival compared to placebo in second line with a HR for OS of 1.05[20]. Data in phase II already suggested a modest median survival of 8.4 months[46] associated to marginal response rates. Also, the companion biomarker study for the phase III trial failed to find any robust predictive biomarker of response to everolimus[47]. Altogether, these data suggest that MTOR inhibition has no anti-tumoral activity in advanced HCC. Other drugs potentially falling under this category include the EGFR inhibitor erlotinib[12], and the FGFR2/VEGFR inhibitor brivanib[10,19]. Limited efficacy can also be claimed for Y-90 resin microsphere treatment as a superior alternative to sorafenib. The phase III trials reported[14,15], were negative with a HR for OS of 1.1. It is important to note that failure to demonstrate superiority does not mean similar efficacy, which requires an *ad hoc* trial design for non-inferiority or equivalence, a concept that will be further discussed in this review[48].

Drug toxicity

The second reason is significant drug toxicity, which is relevant in cirrhotic patients since liver dysfunction decreases the threshold for severe adverse effects. The best example is sunitinib[49], that despite having a molecular target profile similar to sorafenib, it was unable to improve survival when compared in frontline[9]. The trial was prematurely terminated due to futility and safety concerns affecting sunitinib. Median OS for sunitinib was 7.9 months, compared to the 10.2 with sorafenib (HR of 1.3). Treatment-related deaths occurred in 3.2% and 0.4% of patients receiving sunitinib and sorafenib, respectively. When this trial was conducted, sunitinib was already FDA approved for advanced kidney cancer and gastrointestinal stromal tumors, where toxicity was not a major clinical issue. However, in patients with underlying liver disease the toxicity of sunitinib was severe enough to obscure any beneficial anti-tumoral efficacy. Sunitinib has a higher inhibitory potency than sorafenib, particularly regarding its anti-angiogenic activity via VEGFR and PDGFR Inhibition[50]. Angiogenesis is critical during liver fibrogenesis[51], so the strong and sustained anti-angiogenic effect achieved with sunitinib seems detrimental, favoring liver failure. Previous phase II trials testing sunitinib in HCC offer additional insights into the hepatic toxicity of this drug[52–54], including up to 4/37 (11%) treatment-related deaths[54]. An adequate identification of toxicity signals at this stage could help mitigate this problem. Another example is the VEGFR/PDGFR inhibitor linifanib, tested in frontline versus sorafenib[11]. This trial was early terminated based on futility (median OS for linifanib and sorafenib were 9.1 and 9.8 months), but grade 3-4 adverse events were significantly more frequent in linifanib than in sorafenib, including hypertension (21% vs

11%) and hepatic encephalopathy (7% vs 3%). Besides the negative effect of toxicity in clinical outcomes, there is a subtler effect of non-lethal toxicity as it associates with dose reductions, which could also decrease antitumor potency.

Lack of trial enrichment strategies

A third reason for clinical trial failure is lack of effective enrichment strategies for patient enrollment based on predicted biomarkers of response. Trial enrichment in oncology is closely linked to the concept of oncogene addiction. This term describes those molecular alterations, generally DNA mutations or chromosomal aberrations, required for cancer cell proliferation and survival[55]. There are numerous examples in oncology of survival benefits after a clinical trial testing a drug only in those patients with mutations in its target[56] (e.g., ALK rearrangements in lung cancer and response to crizotinib[57]). Only 2/21 (10%) phase III trials in advanced HCC incorporated patient enrichment, likely due to: a) limited access to tumor tissue in patients already diagnosed by noninvasive criteria; and b) few druggable targets among the most common genetic alterations in HCC[58,59]. In fact, the most common mutations in HCC (*TERT* promoter, *CTNNB1*, *TP53*, *AXIN1*, *ARID1A* and *ARID1B*) are untargetable[2,60]. One of them evaluated tivantinib versus placebo in second line in patients with high expression of MET assessed with immunohistochemistry[23]. This trial was based on a post-hoc analysis of 37 patients from a previous phase 2 trial[61] and failed to meet its primary endpoint with a HR for OS of 0.97. Arguably, the signal in the phase 2 trial was weak, but most important, recent data questions the specificity of tivantinib as a MET inhibitor[62]. It was also thought that MET was a prognostic factor, but the median survival of 9.1 months for the placebo arm in MET-high patients in second-line challenges this concept[23]. The second trial tested the VEGFR2 monoclonal antibody ramucirumab versus placebo in second line (i.e. REACH-2) in patients with AFP higher than 400 ng/mL, and showed a significant improvement in OS versus placebo (HR of 0.71[25]). A difference with the tivantinib case is that the rationale for REACH-2 came from a *post hoc* analysis of the negative phase III trial in all-comers (i.e., REACH[21]) which enrolled 565 patients. This showed a robust p of interaction favoring ramucirumab in patients with high AFP of 0.02. AFP is a well-known poor prognostic marker[36,47], highly expressed in tumors with a supposed progenitor cell origin[63], but it does not provide a neat link between any specific driver oncogenic event (i.e., structural DNA alteration or signaling pathway) and ramucirumab's main molecular target. Experimental evidence identifies VEGFR2 as a marker of hepatic progenitors[64], which could hypothetically explain the efficacy of ramucirumab in tumors with high AFP.

Positive phase III clinical trials

Successful drugs in frontline include sorafenib and lenvatinib, whereas regorafenib, cabozantinib, and ramucirumab in patients with high AFP demonstrated efficacy in second line (Table 1 and Figure 2). The PD1 inhibitor nivolumab has shown promising results in phase II with an

ORR of 14% by RECIST (18% by mRECIST) and a median OS of 15.6 months [38,65], which granted its accelerated approval by the FDA. Another immune-based therapy -pembrolizumab- reported similar ORR (17%) but lower median OS (12.9 months[40]). Data from phase III trials in first-line (nivolumab vs sorafenib) and in second-line (pembrolizumab vs placebo) will be critical to fully recommend these immune-based therapies in clinical practice guidelines[4]. Since the strength of evidence so far comes from phase II data, current EASL guidelines just posed a weak recommendation for nivolumab[4].

The reasons for trial failure provide the best clues for the qualities of a drug to be successful in HCC, which essentially are: a) adequate clinical trial design with an emphasis on selection criteria and robust endpoints; b) a fine balance between drug efficacy and toxicity; and c) a proper interpretation of efficacy and toxicity signals in phase II trials. Sorafenib epitomizes these qualities, and to certain extent, the design principles implemented in the pivotal SHARP trial[7] were adopted as best-practices for design in subsequent studies[33]. The target population must include patients with well-preserved liver function (i.e., Child-Pugh A with compensated liver disease) to avoid competing risks from deaths due to progression of the liver disease, and to minimize drug toxicity. Also, patients need to be fit enough to tolerate the drug and with a life expectancy of at least 3 months, which can be reasonably guaranteed by enrolling patients with ECOG performance status test (PST) of 0-1. It is paramount to enroll patients at the same clinical stage as per the BCLC classification[4,66]. The SHARP trial was instrumental to eradicate the misleading concept of 'unresectable' HCC when conducting HCC trials. This concept included a heterogeneous population of patients at intermediate (BCLC-B) and advanced (BCLC-B) stages[67], which imposed significant bias when interpreting trial results. In addition to the same clinical stage, patients need to be adequately stratified for known HCC prognostic factors and geographic region.

Regarding patient's selection, the success of the REACH-2 trial underscores the importance of properly interpreting *post hoc* analysis. The pooled analysis of REACH and REACH-2 assessing ramucirumab in those patients with AFP > 400 ng/mL further confirms a significant and clinically meaningful benefit of ramucirumab vs placebo in second-line (median survival 8.1 vs 5 months; HR=0.694)[25]. The rationale to enrich trials based on predicted oncogene addiction is two-fold: first, to maximize anti-tumoral response by perturbing the cancer drivers active in a given patient, and second, to spare unnecessary toxicity in those patients without the oncogene addiction. HCC has few druggable targets among the most frequent driver mutations, but a recent proof-of-concept trial reinforces the validity of this approach to explore treatment response[68]. A screening of 1,318 HCC patients allowed enrolling 54 cases with RAS mutations (4.4%) detected using circulating tumor DNA (ctDNA) in a phase II trial testing the combination of sorafenib with the MEK inhibitor refametinib[68]. Mutation analysis of ctDNA is feasible in HCC[69] and facilitates screening of large populations. Other potential druggable oncogenic alterations in HCC include high-level DNA amplifications of *FGF19*[70,71] or *VEGFA*[70,72]. Phase II clinical trials are currently exploring

selective inhibition of these candidate oncogene addition loops[73]. There is also increasing interest in developing biomarkers to identify the 20% of patients who respond to immune-based therapies, who show outstanding OS. The use of PD-L1 staining seems irrelevant in HCC[38], and other potential biomarkers such as tumor mutational burden (TMB)[74] or the HCC immune class[75] are under investigation. To facilitate the implementation of biomarker-based clinical trials in HCC, it is essential to enforce mandatory tissue collection in all clinical trials testing new compounds[4]. In this regard, the impact of intratumor heterogeneity in single-biopsy predictions is still debated despite recent studies have shown that driver gene mutations are common between different regions of the tumor[76,77].

Traditionally, new therapies were compared with standard of care or placebo to demonstrate greater efficacy of the new drug. Despite this is the recommended trial design practice in HCC[4,33], some studies after SHARP used non-inferiority designs to challenge sorafenib in first-line. The hypothesis in non-inferiority trials is that the new compound is not substantially worse than the current standard, as opposed to equivalence trials, which are designed to demonstrate that the experimental treatment is neither worse not better than the standard therapy[78]. Non-inferiority trials are required to claim similar efficacy as opposed to assuming it from a negative superiority trial, as previously explained for the Y-90 trials. The non-inferiority trial scenario in HCC is extensively described elsewhere[42], and caveats include the need for larger sample sizes and a very small window of opportunity, as defined by the tight non-inferiority margins. For instance, the BRISK-FL trial was designed to demonstrate non-inferiority of brivanib compared to sorafenib in first line[10]. The trial assumptions set the upper limit of the 95% confidence interval of HR for OS to 1.08. To call non-inferiority, the HR could cross 1, but the upper boundary needed to fall between 1 and 1.08. This threshold is very stringent and can be interpreted as the requirement to demonstrate a robust non-significant trend towards superiority for the new drug. The value proposed by FDA has been calculated based upon capturing at least >60% of the survival benefit obtained with sorafenib[13]. The BRISK-FL trial did not meet this endpoint since the HR confidence interval limits for OS were 0.94 and 1.23. The concept of non-inferiority trials introduces other considerations in treatment recommendations such as toxicity or cost, which will surely contribute to frame the landscape of systemic therapies in HCC.

Hard and surrogate endpoints: Implications in clinical trial design

The overreaching goal of oncological treatments is to allow patients to live longer and better lives than they would do without treatment[79]. Thus, clinical research needs to unequivocally demonstrate statistically and clinically meaningful improvements of the experimental arm over the standard of care. Three types of endpoints have been defined: 1. Hard endpoints, such as overall survival and cancer-specific survival; 2. Surrogate endpoints such as PFS, TTP and ORR, and 3. Patient reported endpoints, such as quality of life (QOL).

Overall survival

This hard endpoint quantifies the time between random trial allocation and death, whatever the cause. Since is not subject to investigator bias, OS has been traditionally recommended by international HCC guidelines as the primary endpoint for randomized phase III trials testing new therapies[80]. In fact, all regular FDA drug approvals in advanced HCC were based upon improvements in OS [7,22]. Cancer-specific survival, where only deaths due to cancer are considered and noncancer-related deaths are censored, is more difficult to assess in conventional trial settings. Deaths due to competing risks, such as liver failure, require a subjective interpretation by the investigator, and thus are more prone to bias[33].

What is the magnitude of benefit to define it as clinically meaningful? This is a controversial topic, highly dependent on the expected outcome for the target population, with confronted opinions between patients, providers, regulatory agencies and health insurers [81,82]. In HCC, there is no consensus on what absolute survival benefit (or the magnitude of benefit in OS as per HR) to be defined as clinically relevant. Reported thresholds of OS with HR<0.8 are also sound for capturing the benefit of patients in advanced HCC trials[83]. ~~An unwritten rule among experts estimate that in advanced HCC scenario, where natural outcome (placebo arm) is estimated to be of around 8 months, absolute gains beyond 2 months are considered clinically relevant, while those below 1 months are not.~~

Survival has some limitations as a sole endpoint in cancer research. First, it might require a long follow-up time to capture enough events due to significant improvement in median OS in the experimental arm[35]. This negatively impacts feasibility and delays access to patients to highly effective drugs. Second, it can be affected by sequential therapies received after tumor progression (post-progression survival), such as for instance regorafenib after a first-line therapy. This might involve one third of patients in recent phase III trials[13]. In this context, validation of surrogate endpoints of OS is paramount to facilitate trial execution and favor a quick deployment of effective drugs in routine clinical conditions.

Surrogate endpoints: PFS, TTP and ORR

Ideally, significant improvement in OS is preferred, but many drugs have been approved based on their ability to improve other less robust endpoints, termed surrogates (i.e., TTP, PFS and ORR). These are outcomes not inherently meaningful from the clinical standpoint, but thought to accurately predict hard outcomes such as OS[84]. The development of surrogate endpoints became a necessity in clinical trials in cardiology, where the long time to accumulate enough events for a hard endpoint made most studies unfeasible. Use of surrogate endpoints is becoming a need in oncology where effective post-progression therapies are available.

Accelerated approval based upon surrogate endpoints is becoming the most relevant path for cancer drug regulatory approval in the US. Between 2009 and 2014; the FDA approved 83 drugs in oncology, 66% of them on the basis of surrogate endpoints[85]. The FDA's accelerated approval program was introduced in 1992 as a social compromise during the worse years of the HIV epidemic to expedite access to agents for life-threating conditions based on surrogate endpoints. The program included a "safety net" that required the manufacturer to conduct post-marketing studies and confirm the efficacy of the drug using hard endpoints[86]. A recent analysis of approved drugs during the period 1992-2017 led to the following conclusions[28]: a) Accelerated approval was granted for 93 indications, ORR being the most common surrogate endpoint used (87% of cases), b) Among drugs approved through this path, 55% were ultimately confirmed for regular approval, 5% of indications were withdrawn (e.g., bevacizumab in metastatic breast cancer[87]), whereas in others the process has not been concluded.

Despite the increasing importance of surrogate endpoints in oncology, they have two main limitations. First, since they usually rely on the radiological definition of tumor progression or response, they are vulnerable to interpretation bias. This can be minimized by using central radiology reviews and a designated adjudicator of response. Second, and more important, in order to be reliable, they require validation as credible predictors of OS[84]. Validation of surrogate endpoints can be conducted at the individual- or trial-levels[88]. While validation of individual-level surrogacy requires individual patient data from at least one clinical trial, trial-level surrogacy uses assembled data from multiple trials. The Institute for Quality and Efficiency in Health Care has proposed a set of criteria to quantify the association between a surrogate and hard endpoint, which includes low (R<0.7), moderate (R>0.7 to R<0.85) and high correlation (R>0.85)[89]. R refers to the weighted Pearson coefficient between HR of OS and HR of the surrogate endpoint. Alternative methods to study this correlation have been reported[90]. A systematic review and meta-analysis of trial-level surrogate endpoints (PFS, TTP and ORR) for OS in oncology including 36 articles and 352 clinical trials found low, moderate and high correlation with OS in 52%, 25% and 23% of surrogate endpoints, respectively[84].

In order to explore the concept of surrogate endpoints recapitulating OS in advanced HCC we have identified 21 RCTs assessing systemic therapies with or without loco-regional therapies in advanced HCC (12 in first-line and 9 in second-line) (Table 1) published between 2008 and 2018 through a MEDLINE search via PubMed using the keywords "advanced hepatocellular carcinoma". Results were limited to "clinical trial, phase III". Trials recently presented at international meetings (2016-2018) were also included despite the full manuscript is not yet available. For each trial, data on sample size, radiological response, TTP and OS were collected. TTP and OS were determined in terms of hazard ratio (HR) using published data (values less than 1 denotes a favourable result in the experimental group). In addition, ORR was established with odds ratio calculated from the published radiological response (values greater than 1 denotes a favourable result in the

experimental group). For the purpose of the trial-level analysis, we first assessed the overall correlation between PFS and OS (R=0.84; R^2=0.71) (Figure 3A), and then the correlation of TTP and OS (R=0.83; R^2=0.69) (Figure 3B). Afterwards, we established a conservative threshold of positive predictive value for PFS since this is the most documented surrogate time-to-event endpoint in oncology, and the one showing a higher correlation with OS[84]. Finally, we explored the correlation between ORR assessed as per mRECIST and survival in early, intermediate and advanced HCC.

Progression-free survival

PFS is a composite endpoint of two variables: death and evidence of radiological progression, usually defined by standard criteria as RECIST[91] or mRECIST[92]. International guidelines initially discouraged this endpoint in HCC due to the competing risk of dying due to progressed liver dysfunction despite a relevant anti-tumoral benefit[33]. However, this limitation has been mitigated since most trials in HCC have adopted restrictive inclusion criteria in terms of liver function (i.e., Child-Pugh A without decompensation). In this scenario, the likelihood of death as a result of liver decompensation (i.e., gastrointestinal bleeding, encephalopathy or ascites and spontaneous peritonitis) is 5% at one year[93]. When we evaluate the association between PFS and OS in HCC phase III trials, we observe a moderate Pearson correlation (R) of 0.84 (Figure 3A). This figure falls in the upper boundary of a moderate correlation (R between 0.7 and 0.85). When specifically analyzing the positive predictive value of theoretical thresholds of PFS correlating with OS, only 3/7 PFS reported a HR ≤0.6 that was significantly associated with a positive survival clinical benefit (in all cases with a HR for OS < 0.8). Conversely, those four studies reporting a positive PFS with a HR between 0.6-0.8 were associated with no significant survival benefits (HR for OS between 0.87 and 1.05) (Figure 3A). In our study, according to the linear regression equation obtained [log HR_{OS} = 0.072 + 0.487 x log HR_{PFS}], a threshold of PFS HR=0.6 (representing a 40% risk reduction) will decrease ~17% the risk of OS (OS HR=0.83) (see Figure 3A). In summary, a moderate correlation has been established between PFS and OS in 21 RCTs in advanced HCC. A value of HR ≤0.6 is proposed as surrogate threshold effect [94], and is likely to predict a clinically meaningful improvement in OS. Is worth to mention though that this rule is supported mainly by positive trials in the second-line setting comparing active drugs vs placebo. We assume that such association is retained in front-line comparing two active drugs, but recommendation in that setting should be tempered due to the lack of confirmatory data.

Time to progression

This endpoint quantifies the time between trial allocation and radiological progression, usually defined by standard criteria as RECIST[91] or mRECIST[92]. Deaths are censored as non-radiological progressions at the time of death or at an earlier visit, with a cause-specific hazard, representing informative censoring. Symmetric repeated radiological measurements every 6-8

weeks are required to avoid missing moderate differences between treatment groups[33]. This recommendation was not followed in the SIRveNIB[15], SARAH[14] and ADI-PEG 20[26] trials.

To delineate the adequacy of TTP as a surrogate of OS in HCC we conducted also a trial-level meta-analysis to evaluate the correlation between TTP and OS in 21 RCTs (Figure 3B). The Pearson correlation (R) was 0.83, which indicates a moderate association according to the IQWIG guidelines[84]. In 10 phase III trials there was a significant difference in TTP in favor of the investigational arm. However, these positive results in TTP were not followed by superiority in OS in 5 (50%) trials. Brivanib[19] and ramucirumab[21] in second-line showed efficacy as per TTP (HR=0.56 and 0.59, respectively), while not significantly improving the hard endpoint of OS (HR=0.89 and 0.87, respectively). Lenvatinib[13] in first-line versus sorafenib also showed significant differences in TTP favoring lenvatinib (HR 0.63), without showing superiority for OS (HR=0.92). This trial was positive since it was designed for non-inferiority (upper 95% CI lower than 1.08). Finally, linifanib[11] and hepatic infusion arterial chemotherapy (HAIC)[18] in first-line failed to show any benefit in terms of OS (HR=1.05 and 1.01, respectively) even though there was a clear benefit when measuring TTP (HR=0.76 and 0.65, respectively). These results do not support the initial recommendation after the SHARP trial of using TTP as the optimal surrogate endpoint[33] in phase 2 trials, and reinforce the need for accurate evaluation of surrogacy in clinical trials. Based on the linear regression model obtained [log HR_{OS} = 0.083 + 0.491 x log HR_{TTP}], we can extrapolate that a therapy producing a 40% risk reduction in TTP will yield an estimated ~16% risk reduction in OS (HR=0.84) (see Figure 3B). Moreover, in order to directly compare the performance of PFS and TTP, we analyzed the correlation between both surrogate endpoints, obtaining a Pearson correlation (R) of 0.99 (Figure 4). Thus, in the modern era of HCC trial design, with minimal cirrhosis-related deaths (due to the inclusion of Child-Pugh A), there is a strong correlation between both endpoints. In fact, when we inferred the non-reported PFS HR of SHARP and AP trials according to the linear equation obtained comparing both surrogate endpoints [log HR_{PFS} = 0.014 + 0.927 x log HR_{TTP}], the HRs values are close to 0.60, just at the previously proposed minimum threshold.

There are two other considerations regarding this endpoint. First, not all types of tumor progression may have the same clinical meaning. Recent data also suggest that TTP may capture heterogeneous features, with essentially two types of progression at advanced stages[95,96]. In particular, survival after progression is significantly worse for patients who develop a new extrahepatic lesion and/or vascular invasion (median OS = 7.1 months) compared to those who progress due to the growth of existing intrahepatic/extrahepatic lesions or the development of a new intrahepatic lesion (median OS = 14.9 months). Second, factors including evaluation bias, trial attrition or informative censoring may weaken the association between the TTP and OS [29]. Finally, prolonged exposure to a given therapy might lead to a phenotypic change in tumors, thus, offsetting any initial advantage from the treatment captured by the surrogate endpoint[97].

Objective response rate

Tumor response in oncology trials is typically measured using the Response Evaluation Criteria in Solid Tumors (RECIST)[91]. These criteria standardize methods for converting radiological observations into a quantitative and statistically tractable framework to define tumor response (i.e., a 30% decrease in the diameter of target lesion). ORR is the percentage of patients who achieve an objective tumor response. Disease control rate (DCR) is the combination of ORR and stable disease, but it has two disadvantages that limit its adoption for regulatory approval: a) the definition of duration of stable disease varies between studies; and b) stable disease can reflect inherent characteristics of the tumor rather than treatment efficacy.

The RECIST criteria were originally developed to evaluate cytotoxic agents. The generalization of targeted therapies has challenged this simplistic approach that relies on tumor shrinkage to indicate clinical efficacy. Sorafenib was associated with only 2-3% of ORR, despite providing clear survival benefits[7,8]. Given the poor correlation between tumor response assessed with conventional tools and OS, a group of experts convened by the American Association for the Study of Liver Diseases (AASLD) proposed specific amendments to standard RECIST[33]. Further description of response and progression resulted in the criteria named modified RECIST (mRECIST), which ultimately incorporates the concept of viable tumor defined as the portions of tumor showing arterial enhancement[92]. The mRECIST criteria in HCC have improved the sensitivity to quantify tumor response with targeted therapies: ORR of 9-17% with sorafenib[10,13,18], 10-12% with brivanib[10,19], 11% with regorafenib[22] and 24% with lenvatinib[13]. Retrospective studies have consistently demonstrated that patients who achieved an objective response on sorafenib had a longer survival than non-responders[98–100]. Recently, data from double-blind randomized trials assessing brivanib and nintedanib further validated this association [101,102]. Thus, the association between tumor response and improved OS in HCC patients at advanced stages complement what was already knew in patients at early and intermediate stages treated with loco-regional therapies[103–109](Table 2).

When we evaluate the trial-level correlation between ORR and OS (Supplementary Figure 1), the R weighted Pearson coefficient obtained is 0.54. This is significantly lower than the correlation obtained with PFS/TTP and OS as depicted in Figure 3. There are two reasons for this: one is inherent to the use of odds ratio instead of hazard ratio to compare differences in ORR. The accuracy of odds ratio decreases for low values of ORRs. The second reason is that only a small proportion of patients within these trials achieved ORR (~10-20%), which is, in fact, the event that correlates with better survival[101]. A direct comparison between RECIST and mRECIST for OS surrogacy through an independent meta-analysis of trials using either criteria would be ideal to define the role of ORR to predict OS in advanced HCC. This will also help determine the best tool to evaluate tumor response to systemic therapies. However, since only 5 RCTs reported response

data using mRECIST, we did not sub-analyze this endpoint according to the tool used to evaluate response.

Some other questions remain unanswered. As observed in other solid tumors treated with efficacious targeted therapies[110–112], the reported rates of responders are still suboptimal to estimate the maximum number of patients who would benefit from the treatment. In addition, the duration of response might be more clinically relevant than the extent of tumor reduction. Finally, the strategy to evaluate response might require a thoughtful revision when assessing immunotherapies. As shown in melanoma patients treated with checkpoint inhibitors, standard RECIST may not provide a reliable assessment of antitumor efficacy[113]. In fact, response to immunotherapy may take longer compared to other agents and can even falsely mirror criteria for progression (i.e., pseudo-progression)[114]. Immune-related response criteria have been developed[115,116], including the concept of "confirmation of progression" by a second scan obtained at least 4 weeks after progressive disease has been registered.

Despite all the challenges that evaluation of tumor response face in oncology, and particularly in HCC, the importance of ORR as a surrogate endpoint is recognized by regulatory agencies and frequently used for accelerated drug approval. This was the case of nivolumab, approved in second-line based on an ORR of 18% by mRECIST and 14% by RECIST [38,39]. Remarkably, objective response to nivolumab has been associated with prolonged OS[65]. Overall, the fact that a high ORR in phase II trials was considered a robust criterion for drug approval[117], and further success in phase III trials[118], indicates that ORR should be considered as a primary endpoint for single-arm phase II studies. Related to this, early clinical trials are showing promising results with combinations of checkpoint inhibitors and targeted therapies, as measured by ORR. Lenvatinib plus pembrolizumab[119] and atezolizumab plus bevacizumab[120] achieved an ORR of 46% by mRECIST and 50% by RECIST in advanced HCC, respectively. As a result, the later combination was granted breakthrough therapy designation by the FDA[121]. Of note, most of the drugs approved under the accelerated program reported ORR exceeding 30%[122].

Patient-reported endpoints: Quality of life

Health-related QOL measures the effect of the disease on an individual's physical, psychological and social functioning and well-being[123]. Regulatory agencies recognize symptomatic improvement as a direct clinical benefit to patients and an important consideration in drug approval[124]. However, unlike OS, the interpretation of QOL is subjective. In HCC, two tools have been proposed to measure QOL: the European Organization for Research and Treatment of Cancer Quality of Live Questionnaire (EORTC QLQ-HCC18)[125] and the Functional Assessment of Cancer Therapy-Hepatobiliary (FACT-Hep) questionnaire[126]. They can be used to evaluate time to symptomatic progression (i.e. time between trial allocation and the occurrence of disease related symptoms according to preestablished scores). QOL was measured in the SHARP trial[7]

according to the FHSI-8 questionnaire[127], a reduced version of FACT-Hep, with results that collided with OS. Conversely, the SARAH trial[14] reported better global health status with Y-90 when compared to sorafenib based on QLQ-HCC18, which was inconsistent with the primary endpoint of OS. Evaluation of QOL is contingent on when it was assessed during disease progression. Also, significant changes in QOL have been observed across different cultures[128]. Defining and evaluating reliable QOL assessment tools has been established as one of the unmet needs in HCC research by international guidelines[4]. In summary, health-related QOL measures are not ready to support, as single tools, regulatory approval for drugs in HCC.

Conclusions

The current period of drug development in HCC is providing major advancements in the management of this devastating disease. Six drugs have currently shown activity as systemic therapies, which represents an unprecedented revolution for the last 50 years. Novel drugs or combinations strategies are emerging in the field, and thus new tools will be required for the proper assessment of clinical benefits. OS is still the most robust endpoint but the increasing number of treatments available in advanced HCC preludes the use of surrogate endpoints, less vulnerable to subsequent treatments after progression. In this scenario, PFS has shown moderate correlation with OS (R=0.84), and a threshold of HR≤0.6 defines a conservative approach of surrogate endpoint able to capture survival differences in a superiority trial with a high positive predictive value. Two recent studies have been released supporting our threshold of HR ≤ 0.6 for PFS. The first one, an individual-patient data meta-analysis of two RCTs (REACH[21] and REACH-2[25]), showing a significant OS HR with a PFS HR of 0.57[129]. The second one, a phase 3 RCT comparing sorafenib with or without cTACE in advanced HCC, a negative study for its primary endpoint (OS) with a PFS HR of 0.73[130]. Thus, PFS-HR ≤0.6 could be considered a candidate endpoint in phase II and phase III RCTs when subsequent therapies are expected to impact overall outcome. ORR by sensitive criteria (mRECIST) may be useful particularly in single arm phase II trials with proof of concept drugs or in combination studies targeting accelerated approval with a threshold >30%. Finally, the current development of RCTs assessing immune therapies or drug combinations in HCC will certainly evolve the paradigm of drug development and trial design. Particularly of interest will be whether the statements proposed in the present review are confirmed in trials designed with composite primary endpoints, such as OS-PFS for lenvatinib+pembrolizumab vs lenvatinib (NCT03713593) or OS-ORR for atezolizumab+bevacizumab vs sorafenib (NCT03434379).

Key points

- In the last two years four systemic agents (i.e., regorafenib, lenvatinib, cabozantinib and ramucirumab) have shown clinical benefit in the setting of phase III trials and one (i.e.,

nivolumab) has been granted accelerated approval based on a phase II trial, expanding, thus, the pipeline of effective drugs available in advanced HCC to providers.

- The improvement in the number of effective agents comes at a cost of increased complexity of clinical decision-making, and thus, in the design of future clinical trials.
- OS is still the most robust endpoint in advanced HCC but the increasing number of treatments after progression underscore the need for surrogate endpoints.
- PFS has a moderate correlation at trial level with OS (R=0.84). A conservative minimum surrogate threshold effect of HR ≤0.6 is highly predictive of a significant improvement in OS, whereas HR ranging from 0.6 – 0.7 are uncertain surrogates.
- ORR by sensitive criteria in single arm phase II trials could be a useful tool to prioritize treatments for testing in phase III trials.

References

[1] Torre L, Bray F, Siegel R, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA Cancer J Clin 2015;65:87–108.

[2] Llovet JM, Zucman-Rossi J, Pikarsky E, Sangro B, Schwartz M, Sherman M, et al. Hepatocellular carcinoma. Nat Rev Dis Prim 2016;2:16018.

[3] Llovet JM, Brú C, Bruix J. Prognosis of hepatocellular carcinoma: the BCLC staging classification. Semin Liver Dis 1999;19:329–38.

[4] Galle PR, Forner A, Llovet JM, Mazzaferro V, Piscaglia F, Raoul J, et al. EASL Clinical Practice Guidelines: Management of hepatocellular carcinoma. J Hepatol 2018;69:182–236.

[5] Llovet JM, Bruix J. Systematic review of randomized trials for unresectable hepatocellular carcinoma: Chemoembolization improves survival. Hepatology 2003;37:429–42.

[6] Llovet JM, Montal R, Sia D, Finn RS. Molecular therapies and precision medicine for hepatocellular carcinoma. Nat Rev Clin Oncol 2018;In press.

[7] Llovet J, Ricci S, Mazzaferro V, Hilgard P, Gane E, Blanc J, et al. Sorafenib in advanced hepatocellular carcinoma. N Engl J Med 2008;359:378–90.

[8] Cheng A-L, Kang Y-K, Chen Z, Tsao C-J, Qin S, Kim JS, et al. Efficacy and safety of sorafenib in patients in the Asia-Pacific region with advanced hepatocellular carcinoma: a phase III randomised, double-blind, placebo-controlled trial. Lancet Oncol 2009;10:25–34.

[9] Cheng A-L, Kang Y-K, Lin D-Y, Park J-W, Kudo M, Qin S, et al. Sunitinib versus sorafenib in advanced hepatocellular cancer: results of a randomized phase III trial. J Clin Oncol 2013;31:4067–75.

[10] Johnson PJ, Qin S, Park J-W, Poon RTP, Raoul J-L, Philip PA, et al. Brivanib versus

sorafenib as first-line therapy in patients with unresectable, advanced hepatocellular carcinoma: results from the randomized phase III BRISK-FL study. J Clin Oncol 2013;31:3517–24.

[11] Cainap C, Qin S, Huang W-T, Chung IJ, Pan H, Cheng Y, et al. Linifanib versus Sorafenib in patients with advanced hepatocellular carcinoma: results of a randomized phase III trial. J Clin Oncol 2015;33:172–9.

[12] Zhu a. X, Rosmorduc O, Evans TRJ, Ross PJ, Santoro A, Carrilho FJ, et al. SEARCH: A Phase III, Randomized, Double-Blind, Placebo-Controlled Trial of Sorafenib Plus Erlotinib in Patients With Advanced Hepatocellular Carcinoma. J Clin Oncol 2014;33:559–66.

[13] Kudo M, Finn R, Qin S, Han SS, Ikeda K, Piscaglia F, et al. A Randomised Phase 3 Trial of Lenvatinib vs. Sorafenib in Firstline Treatment of Patients With Unresectable Hepatocellular Carcinoma. Lancet 2018;391:1163–73.

[14] Vilgrain V, Pereira H, Assenat E, Guiu B, Ilonca AD, Pageaux GP, et al. Efficacy and safety of selective internal radiotherapy with yttrium-90 resin microspheres compared with sorafenib in locally advanced and inoperable hepatocellular carcinoma (SARAH): an open-label randomised controlled phase 3 trial. Lancet Oncol 2017;18:1624–36.

[15] Chow PKH, Gandhi M, Tan S-B, Khin MW, Khasbazar A, Ong J, et al. SIRveNIB: Selective Internal Radiation Therapy Versus Sorafenib in Asia-Pacific Patients With Hepatocellular Carcinoma. J Clin Oncol 2018;36:1913–21.

[16] Qin S, Bai Y, Lim HY, Thongprasert S, Chao Y, Fan J, et al. Randomized, multicenter, open-label study of oxaliplatin plus fluorouracil/leucovorin versus doxorubicin as palliative chemotherapy in patients with advanced hepatocellular carcinoma from Asia. J Clin Oncol 2013;31:3501–8.

[17] Abou-Alfa GK, Niedzwieski D, Knox JJ, Kaubisch A, Posey J, Tan BR, et al. Phase III randomized study of sorafenib plus doxorubicin versus sorafenib in patients with advanced hepatocellular carcinoma (HCC): CALGB 80802 (Alliance). J Clin Oncol 2016;34:4_suppl.

[18] Kudo M, Ueshima K, Yokosuka O, Ogasawara S, Obi S, Izumi N, et al. Sorafenib plus low-dose cisplatin and fluorouracil hepatic arterial infusion chemotherapy versus sorafenib alone in patients with advanced hepatocellular carcinoma (SILIUS): a randomised, open label, phase 3 trial. Lancet Gastroenterol Hepatol 2018;3:424–32.

[19] Llovet JM, Decaens T, Raoul JL, Boucher E, Kudo M, Chang C, et al. Brivanib in patients with advanced hepatocellular carcinoma who were intolerant to sorafenib or for whom sorafenib failed: results from the randomized phase III BRISK-PS study. J Clin Oncol 2013;31:3509–16.

[20] Zhu AX, Kudo M, Assenat E, Cattan S, Kang Y-K, Lim HY, et al. Effect of everolimus on survival in advanced hepatocellular carcinoma after failure of sorafenib: the EVOLVE-1 randomized clinical trial. JAMA 2014;312:57–67.

[21] Zhu AX, Park JO, Ryoo B-Y, Yen C-J, Poon R, Pastorelli D, et al. Ramucirumab versus placebo as second-line treatment in patients with advanced hepatocellular carcinoma following first-line therapy with sorafenib (REACH): a randomised, double-blind, multicentre, phase 3 trial. Lancet Oncol 2015:859–70.

[22] Bruix J, Qin S, Merle P, Granito A, Huang Y-H, Bodoky G, et al. Regorafenib for patients with hepatocellular carcinoma who progressed on sorafenib treatment (RESORCE): a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet 2017;389:56–66.

[23] Rimassa L, Assenat E, Peck-Radosavljevic M, Zagonel V, Pracht M, Rota Caremolli E. Second-line tivantinib (ARQ 197) vs placebo in patients (Pts) with MET-high hepatocellular carcinoma (HCC): Results of the METIV-HCC phase III trial. J Clin Oncol 2017;35:15_supp.

[24] Abou-Alfa GK, Meyer T, Cheng A-L, El-Khoueiry AB, Rimassa L, Ryoo B-Y, et al. Cabozantinib in Patients with Advanced and Progressing Hepatocellular Carcinoma. N Engl J Med 2018;379:54–63.

[25] Zhu AX, Kang Y-K, Yen C-J, Finn RS, Galle PR, Llovet JM, et al. REACH-2: A randomized, double-blind, placebo-controlled phase 3 study of ramucirumab versus placebo as second-line treatment in patients with advanced hepatocellular carcinoma (HCC) and elevated baseline alpha-fetoprotein (AFP) following first-line sorafe. J Clin Oncol 2018;36 Suppl.

[26] Abou-Alfa GK, Qin S, Lu S-N, Yen C-J, Feng Y-H, Lim Y, et al. Phase III randomized study of second line ADI-peg 20 (A) plus best supportive care versus placebo (P) plus best supportive care in patients (pts) with advanced hepatocellular carcinoma (HCC). Ann Oncol 2018;29:1402–8.

[27] Merle P, Bodoky G, López-López C, Saad AS, Casadei Gardini A, Borbath I, et al. Safety and Efficacy Results from the Phase 3 Relive Study of Doxorubicin-Loaded Nanoparticles versus Best Standard of Care in Patients with Advanced Hepatocellular Carcinoma after Failure or Intolerance to Previous Treatment Including Sorafenib. Int Liver Cancer Assoc 11th Annu Conf 2017.

[28] Beaver JA, Howie LJ, Pelosof L, Kim T, Liu J, Goldberg KB, et al. A 25-Year Experience of US Food and Drug Administration Accelerated Approval of Malignant Hematology and Oncology Drugs and Biologics. JAMA Oncol 2018;4:849–56.

[29] Kemp R, Prasad V. Surrogate endpoints in oncology: When are they acceptable for regulatory and clinical decisions, and are they currently overused? BMC Med 2017;15:1–7.

[30] Bruix J SMAA for the S of LD. Management of hepatocellular carcinoma: an update. Hepatology 2011;53:1020–2.

[31] Yeo W, Mok TS, Zee B, Leung TWT, Lai PBS, Lau WY, et al. A randomized phase III study of doxorubicin versus cisplatin/interferon alpha-2b/doxorubicin/fluorouracil (PIAF) combination chemotherapy for unresectable hepatocellular carcinoma. J Natl Cancer Inst 2005;97:1532–8.

[32] Chow P, Tai B-C, Tan C-K, Machin D, Win KM, Johnson PJ, et al. High-dose tamoxifen in the treatment of inoperable hepatocellular carcinoma: A multicenter randomized controlled trial. Hepatology 2002;36:1221–6.

[33] Llovet JM, Di Bisceglie AM, Bruix J, Kramer BS, Lencioni R, Zhu AX, et al. Design and endpoints of clinical trials in hepatocellular carcinoma. J Natl Cancer Inst 2008;100:698–711.

[34] Bruix J, Cheng A-L, Meinhardt G, Nakajima K, De Sanctis Y, Llovet J. Prognostic Factors and Predictors of Sorafenib Benefit in Patients With Hepatocellular Carcinoma: Analysis of Two Phase 3 Studies. J Hepatol 2017;67:999–1008.

[35] Finn RS, Merle P, Granito A, Huang Y-H, Bodoky G, Pracht M, et al. Outcomes with sorafenib followed by regorafenib or placebo for HCC: additional analyses from the phase 3 RESORCE trial. J Hepatol 2018;In press.

[36] Llovet JM, Pena CE a., Lathia CD, Shan M, Meinhardt G, Bruix J. Plasma Biomarkers as Predictors of Outcome in Patients with Advanced Hepatocellular Carcinoma. Clin Cancer Res 2012;18:2290–300.

[37] Ricke J, Sangro B, Amthauer H, Bargellini I, Bartenstein P, De Toni E, et al. The impact of combining Selective Internal Radiation Therapy (SIRT) with sorafenib on overall survival in patients with advanced hepatocellular carcinoma: The SORAMIC trial palliative cohort. J Hepatol 2018;68 | S65–S.

[38] El-Khoueiry AB, Sangro B, Yau T, Crocenzi TS, Kudo M, Hsu C, et al. Nivolumab in patients with advanced hepatocellular carcinoma (CheckMate 040): An open-label, non-comparative, phase 1/2 dose escalation and expansion trial. Lancet 2017;389:2492–502.

[39] FDA grants accelerated approval to nivolumab for HCC previously treated with sorafenib n.d. https://www.fda.gov/Drugs/InformationOnDrugs/ApprovedDrugs/ucm577166.htm.

[40] Zhu AX, Finn RS, Edeline J, Cattan S, Ogasawara S, Palmer D, et al. Pembrolizumab in patients with advanced hepatocellular carcinoma previously treated with sorafenib (KEYNOTE-224): a non-randomised, open-label phase 2 trial. Lancet Oncol 2018;19.

[41] Gan HK, You B, Pond GR, Chen EX. Assumptions of expected benefits in randomized phase III trials evaluating systemic treatments for cancer. J Natl Cancer Inst 2012;104:590–8.

[42] Llovet JM, Hernandez-Gea V. Hepatocellular carcinoma: reasons for phase III failure and novel perspectives on trial design. Clin Cancer Res 2014;20:2072–9.

[43] Guri Y, Colombi M, Dazert E, Hindupur SK, Roszik J, Moes S, et al. mTORC2 Promotes Tumorigenesis via Lipid Synthesis. Cancer Cell 2017;32:807–823.e12.

[44] Villanueva A, Chiang DY, Newell P, Peix J, Thung S, Alsinet C, et al. Pivotal Role of mTOR Signaling in Hepatocellular Carcinoma. Gastroenterology 2008;135:1972–83.

[45] Sahin F, Kannangai R, Adegbola O, Wang J, Su G, Torbenson M. mTOR and P70 S6 kinase expression in primary liver neoplasms. Clin Cancer Res 2004;10:8421–5.

[46] Zhu AX, Abrams TA, Miksad R, Blaszkowsky LS, Meyerhardt JA, Zheng H, et al. Phase 1/2

study of everolimus in advanced hepatocellular carcinoma. Cancer 2011;117:5094–102.

[47] Zhu AX, Chen D, He W, Kanai M, Voi M, Chen LT, et al. Integrative biomarker analyses indicate etiological variations in hepatocellular carcinoma. J Hepatol 2016;65:296–304.

[48] Llovet JM, Finn RS. Negative phase 3 study of 90Y microspheres versus sorafenib in HCC. Lancet Oncol 2018;19:e69.

[49] Faivre S, Delbaldo C, Vera K, Robert C, Lozahic S, Lassau N, et al. Safety, pharmacokinetic, and antitumor activity of SU11248, a novel oral multitarget tyrosine kinase inhibitor, in patients with cancer. J Clin Oncol 2006;24:25–35.

[50] Stein MN, Flaherty KT. Sorafenib and sunitinib in renal cell carcinoma. Clin Cancer Res 2007;13:3765–70.

[51] Poisson J, Lemoinne S, Boulanger C, Durand F, Moreau R, Valla D, et al. Liver sinusoidal endothelial cells: Physiology and role in liver diseases. J Hepatol 2017;66:212–27.

[52] Koeberle D, Montemurro M, Samaras P, Majno P, Simcock M, Limacher A, et al. Continuous Sunitinib Treatment in Patients with Advanced Hepatocellular Carcinoma: A Swiss Group for Clinical Cancer Research (SAKK) and Swiss Association for the Study of the Liver (SASL) Multicenter Phase II Trial (SAKK 77/06). Oncologist 2010;15:285–92.

[53] Zhu AX, Sahani D V., Duda DG, Di Tomaso E, Ancukiewicz M, Catalano OA, et al. Efficacy, safety, and potential biomarkers of sunitinib monotherapy in advanced hepatocellular carcinoma: A phase II study. J Clin Oncol 2009;27:3027–35.

[54] Faivre S, Raymond E, Boucher E, Douillard J, Lim HY, Kim JS, et al. Safety and efficacy of sunitinib in patients with advanced hepatocellular carcinoma: an open-label, multicentre, phase II study. Lancet Oncol 2009;10:794–800.

[55] Luo J, Solimini NL, Elledge SJ. Principles of Cancer Therapy: Oncogene and Non-oncogene Addiction. Cell 2009;136:823–37.

[56] Villanueva A, Llovet JM. Targeted therapies for hepatocellular carcinoma. Gastroenterology 2011;140:1410–26.

[57] Shaw AT, Kim D-W, Nakagawa K, Seto T, Crinó L, Ahn M-J, et al. Crizotinib versus Chemotherapy in Advanced *ALK* -Positive Lung Cancer. N Engl J Med 2013;368:2385–94.

[58] Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. Nat Med 2017;23:703–13.

[59] Schulze K, Nault J-C, Villanueva A. Genetic profiling of hepatocellular carcinoma using next-generation sequencing. J Hepatol 2016;65:1031–42.

[60] Zucman-Rossi J, Villanueva A, Nault JC, Llovet JM. Genetic Landscape and Biomarkers of Hepatocellular Carcinoma. Gastroenterology 2015;149:1226–39.

[61] Santoro A, Rimassa L, Borbath I, Daniele B, Salvagni S, Van Laethem JL, et al. Tivantinib for second-line treatment of advanced hepatocellular carcinoma: A randomised, placebo-

controlled phase 2 study. Lancet Oncol 2013;14:55–63.

[62] Rebouissou S, La Bella T, Rekik S, Imbeaud S, Calatayud AL, Rohr-Udilova N, et al. Proliferation markers are associated with MET expression in hepatocellular carcinoma and predict tivantinib sensitivity in vitro. Clin Cancer Res 2017;23:4364–75.

[63] Hoshida Y, Nijman SMB, Kobayashi M, Chan JA, Brunet J-P, Chiang DY, et al. Integrative transcriptome analysis reveals common molecular subclasses of human hepatocellular carcinoma. Cancer Res 2009;69:7385–92.

[64] Goldman O, Han S, Sourrisseau M, Dziedzic N, Hamou W, Corneo B, et al. KDR identifies a conserved human and murine hepatic progenitor and instructs early liver development. Cell Stem Cell 2013;12:748–60.

[65] El-Khoueiry AB, Melero I, Yau TC, Crocenzi TS, Kudo M, Hsu C, et al. Impact of antitumor activity on survival outcomes, and nonconventional benefit, with nivolumab (NIVO) in patients with advanced hepatocellular carcinoma (aHCC): Subanalyses of CheckMate-040. J Clin Oncol 2018;36, suppl.

[66] Forner A, Reig M, Bruix J. Hepatocellular carcinoma. Lancet 2018;391:1301–14.

[67] Llovet JM, Bustamante J, Castells a, Vilana R, Ayuso MDC, Sala M, et al. Natural history of untreated nonsurgical hepatocellular carcinoma: rationale for the design and evaluation of therapeutic trials. Hepatology 1999;29:62–7.

[68] Lim HY, Merle P, Weiss KH, Yau T, Ross PJ, Mazzaferro V, et al. Phase II Studies with Refametinib or Refametinib plus Sorafenib in Patients with RAS-mutated Hepatocellular Carcinoma. Clin Cancer Res 2018;June 27.

[69] Labgaa I, Villacorta-Martin C, D'Avola D, Craig AJ, von Felden J, Martins-Filho SN, et al. A pilot study of ultra-deep targeted sequencing of plasma DNA identifies driver mutations in hepatocellular carcinoma. Oncogene 2018:1–13.

[70] Chiang DY, Villanueva A, Hoshida Y, Peix J, Newell P, Minguez B, et al. Focal gains of VEGFA and molecular classification of hepatocellular carcinoma. Cancer Res 2008;68:6779–88.

[71] Sawey ET, Chanrion M, Cai C, Wu G, Zhang J, Zender L, et al. Identification of a therapeutic strategy targeting amplified FGF19 in liver cancer by Oncogenomic screening. Cancer Cell 2011;19:347–58.

[72] Horwitz E, Stein I, Andreozzi M, Nemeth J, Shoham A, Pappo O, et al. Human and mouse VEGFA-amplified hepatocellular carcinomas are highly sensitive to Sorafenib treatment. Cancer Discov 2014;4:730–43.

[73] Kang Y-K, Macarulla T, Yau T, Sarker D, Choo SP, Meyer T, et al. Clinical Activity of Blu-554, a Potent, Highly-Selective FGFR4 Inhibitor in Advanced Hepatocellular Carcinoma (HCC) with FGFR4 Pathway Activation. Int Liver Cancer Assoc 11th Annu Conf 2017.

[74] Hellmann MD, Ciuleanu T-E, Pluzanski A, Lee JS, Otterson GA, Audigier-Valette C, et al.

Nivolumab plus Ipilimumab in Lung Cancer with a High Tumor Mutational Burden. N Engl J Med 2018;378:2093–104.

[75] Sia D, Jiao Y, Martinez-Quetglas I, Kuchuk O, Villacorta-Martin C, Castro De Moura M. Identification of an Immune-specific Class of Hepatocellular Carcinoma, Based on Molecular Features. Gastroenterology 2017;153:812–26.

[76] Torrecilla S, Sia D, Harrington AN, Zhang Z, Cabellos L, Cornella H, et al. Trunk mutational events present minimal intra- and inter-tumoral heterogeneity in hepatocellular carcinoma. J Hepatol 2017;67:1222–31.

[77] Reiter JG, Makohon-Moore AP, Gerold JM, Heyde A, Attiyeh MA, Kohutek ZA, et al. Minimal functional driver gene heterogeneity among untreated metastases. Science 2018;361:1033–7.

[78] Kaji AH, Lewis RJ. Noninferiority Trials: Is a New Treatment Almost as Effective as Another? Jama 2015;313:2371–2.

[79] Wilson MK, Karakasis K, Oza AM. Outcomes and endpoints in trials of cancer treatment: the past, present, and future. Lancet Oncol 2015;16:e32–42.

[80] Llovet JM, Di Bisceglie AM, Bruix J, Kramer BS, Lencioni R, Zhu AX, et al. Design and endpoints of clinical trials in hepatocellular carcinoma. J Natl Cancer Inst 2008;100:698–711.

[81] Booth CM, Tannock I. Reflections on medical oncology: 25 years of clinical trials--where have we come and where are we going? J Clin Oncol 2008;26:6–8.

[82] Cherny NI, Dafni U, Bogaerts J, Latino NJ, Pentheroudakis G, Douillard JY, et al. ESMO-Magnitude of Clinical Benefit Scale version 1.1. Ann Oncol 2017;28:2340–66.

[83] Ellis LM, Bernstein DS, Voest EE, Berlin JD, Sargent D, Cortazar P, et al. American society of clinical oncology perspective: Raising the bar for clinical trials by defining clinically meaningful outcomes. J Clin Oncol 2014;32:1277–80.

[84] Prasad V, Kim C, Burotto M, Vandross A. The strength of association between surrogate end points and survival in oncology: A systematic review of trial-level meta-analyses. JAMA Intern Med 2015;175:1389–98.

[85] Kim C, Prasad V. Strength of Validation for Surrogate End Points Used in the US Food and Drug Administration's Approval of Oncology Drugs. Mayo Clin Proc 2016;91:713–25.

[86] Gyawali B, Kesselheim AS. Reinforcing the social compromise of accelerated approval. Nat Rev Clin Oncol 2018;July 3.

[87] Carpenter D, Kesselheim AS, Joffe S. Reputation and Precedent in the Bevacizumab Decision. N Engl J Med 2011;365:e3.

[88] Zhao F. Surrogate End Points and Their Validation in Oncology Clinical Trials. J Clin Oncol 2016;34:1436–7.

[89] Institute for Quality and Efficiency in Health Care [Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen] (IQWiG). Validity of surrogate endpoints in oncology: executive

summary. n.d. http: //www.iqwig.de/download/A10-05_Executive _Summary_v1-1_Surrogate_endpoints_in_oncology.pdf.

[90] Ciani O, Buyse M, Garside R, Peters J, Saad ED, Stein K, et al. Meta-analyses of randomized controlled trials show suboptimal validity of surrogate outcomes for overall survival in advanced colorectal cancer. J Clin Epidemiol 2015;68:833–42.

[91] Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). Eur J Cancer 2009;45:228–47.

[92] Lencioni R, Llovet J. Modified RECIST (mRECIST) Assessment for Hepatocellular Carcinoma. Semin Liver Dis 2010;30:052–60.

[93] D'Amico G, Garcia-Tsao G, Pagliaro L. Natural history and prognostic indicators of survival in cirrhosis: A systematic review of 118 studies. J Hepatol 2006;44:217–31.

[94] Burzykowski T, Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. Pharm Stat 2006;5:173–86.

[95] Reig M, Rimola J, Torres F, Darnell A, Rodriguez-Lope C, Forner A, et al. Postprogression survival of patients with advanced hepatocellular carcinoma: rationale for second-line trial design. Hepatology 2013;58:2023–31.

[96] Iavarone M, Cabibbo G, Biolato M, Della Corte C, Maida M, Barbara M, et al. Predictors of survival in patients with advanced hepatocellular carcinoma who permanently discontinued sorafenib. Hepatology 2015;62:784–91.

[97] Booth CM, Eisenhauer EA, Group NCT. Progression-free survival: meanigful or simply measurable? J Clin Oncol 2012;30:1030–3.

[98] Edeline J, Boucher E, Rolland Y, Vauléon E, Pracht M, Perrin C, et al. Comparison of tumor response by Response Evaluation Criteria in Solid Tumors (RECIST) and modified RECIST in patients treated with sorafenib for hepatocellular carcinoma. Cancer 2012;118:147–56.

[99] Ronot M, Bouattour M, Wassermann J, Bruno O, Dreyer C, Larroque B, et al. Alternative Response Criteria (Choi, European association for the study of the liver, and modified Response Evaluation Criteria in Solid Tumors [RECIST]) Versus RECIST 1.1 in patients with advanced hepatocellular carcinoma treated with sorafenib. Oncologist 2014;19:394–402.

[100] Takada J, Hidaka H, Nakazawa T, Kondo M, Numata K, Tanaka K, et al. Modified response evaluation criteria in solid tumors is superior to response evaluation criteria in solid tumors for assessment of responses to sorafenib in patients with advanced hepatocellular carcinoma. BMC Res Notes 2015;8:609.

[101] Lencioni R, Montal R, Torres F, Park J-W, Decaens T, Raoul J-L, et al. Objective response by mRECIST as a predictor and potential surrogate end-point of overall survival in advanced HCC. J Hepatol 2017;66:1166–72.

[102] Meyer T, Palmer DH, Cheng A-L, Hocke J, Loembé A-B, Yen C-J. mRECIST to predict

survival in advanced hepatocellular carcinoma: Analysis of two randomised phase II trials comparing nintedanib vs sorafenib. Liver Int 2017:1–9.

[103] Cabibbo G, Maida M, Genco C, Alessi N, Peralta M, Butera G, et al. Survival of Patients with Hepatocellular Carcinoma (HCC) Treated by Percutaneous Radio-Frequency Ablation (RFA) Is Affected by Complete Radiological Response. PLoS One 2013;8.

[104] Vincenzi B, Di Maio M, Silletta M, D'Onofrio L, Spoto C, Piccirillo MC, et al. Prognostic Relevance of Objective Response According to EASL Criteria and mRECIST Criteria in Hepatocellular Carcinoma Patients Treated with Loco-Regional Therapies: A Literature-Based Meta-Analysis. PLoS One 2015;10:e0133488.

[105] Sala M, Llovet JM, Vilana R, Bianchi L, Solé M, Ayuso C, et al. Initial response to percutaneous ablation predicts survival in patients with hepatocellular carcinoma. Hepatology 2004;40:1352–60.

[106] Gillmore R, Stuart S, Kirkwood A, Hameeduddin A, Woodward N, Burroughs AK, et al. EASL and mRECIST responses are independent prognostic factors for survival in hepatocellular cancer patients treated with transarterial embolization. J Hepatol 2011;55:1309–16.

[107] Kim BK, Kim KA, Park JY, Ahn SH, Chon CY, Han KH, et al. Prospective comparison of prognostic values of modified Response Evaluation Criteria in Solid Tumours with European Association for the Study of the Liver criteria in hepatocellular carcinoma following chemoembolisation. Eur J Cancer 2013;49:826–34.

[108] Jung ES, Kim JHJS, Yoon EL, Lee HJ, Lee SJ, Suh SJ, et al. Comparison of the methods for tumor response assessment in patients with hepatocellular carcinoma undergoing transarterial chemoembolization. J Hepatol 2013;58:1181–7.

[109] Prajapati HJ, Spivey JR, Hanish SI, El-rayes BF, Kauh JS, Chen Z, et al. mRECIST and EASL responses at early time point by contrast-enhanced dynamic MRI predict survival in patients with unresectable hepatocellular carcinoma (HCC) treated by doxorubicin drug-eluting beads transarterial chemoembolization (DEB TACE). Ann Oncol 2013;24:965–73.

[110] Hudes G, Carducci M, Tomczak P, Dutcher J, Figlin R, Kapoor A, et al. Temsirolimus, interferon alpha or both for advanced renal cell carcinoma. N Engl J Med 2007;356:2271–81.

[111] Karapetis CS, Khambata-Ford S, Jonker DJ, O'Callaghan CJ, Tu D, Tebbutt NC, et al. K-ras mutations and benefit from cetuximab in advanced colorectal cancer. N Engl J Med 2008;359:1757–65.

[112] Hurwitz H, Fehrenbacher L, Novotny W, Cartwright T, Hainsworth J, Heim W, et al. Bevacizumab plus Irinotecan, Fluorouracil, and Leucovorin for Metastatic Colorectal Cancer. N Engl J Med 2004;350:2335–42.

[113] Hodi FS, Hwu WJ, Kefford R, Weber JS, Daud A, Hamid O, et al. Evaluation of immune-related response criteria and RECIST v1.1 in patients with advanced melanoma treated with Pembrolizumab. J Clin Oncol 2016;34:1510–7.

[114] Wolchok JD, Hoos A, O'Day S, Weber JS, Hamid O, Lebbé C, et al. Guidelines for the evaluation of immune therapy activity in solid tumors: Immune-related response criteria. Clin Cancer Res 2009;15:7412–20.

[115] Seymour L, Bogaerts J, Perrone A, Ford R, Schwartz LH, Mandrekar S, et al. iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics. Lancet Oncol 2017;18:e143–52.

[116] Stephen Hodi F, Ballinger M, Lyons B, Soria JC, Nishino M, Tabernero J, et al. Immune-modified response evaluation criteria in solid tumors (imrecist): Refining guidelines to assess the clinical benefit of cancer immunotherapy. J Clin Oncol 2018;36:850–8.

[117] Tsimberidou AM, Braiteh F, Stewart DJ, Kurzrock R. Ultimate fate of oncology drugs approved by the US food and drug administration without a randomized trial. J Clin Oncol 2009;27:6243–50.

[118] El-Maraghi RH, Eisenhauer EA. Review of Phase II Trial Designs Used in Studies of Molecular Targeted Agents: Outcomes and Predictors of Success in Phase III. J Clin Oncol 2008;26:1346–54.

[119] Ikeda M, Sung MW, Kudo M, Kobayashi M, Baron AD, Finn RS, et al. A phase 1b trial of lenvatinib (LEN) plus pembrolizumab (PEM) in patients (pts) with unresectable hepatocellular carcinoma (uHCC). J Clin Oncol 2018;36, suppl.

[120] Stein S, Pishvaian MJ, Lee MS, Lee K-H, Hernandez S, Kwan A, et al. Safety and clinical activity of 1L atezolizumab + bevacizumab in a phase Ib study in hepatocellular carcinoma (HCC). J Clin Oncol 2018;36, suppl.

[121] FDA Grants Breakthrough Therapy Designation for Genentech's TECENTRIQ in Combination With Avastin as First-Line Treatment for Advanced or Metastatic Hepatocellular Carcinoma (HCC) 2018. https://www.gene.com/media/press-releases/14736/2018-07-17/fda-grants-breakthrough-therapy-designat?utm_source=F&utm_medium=P&utm_term=15538&utm_content=TecentriqHCCBTD&utm_campaign=TecentriqHCCBTD.

[122] Oxnard GR, Wilcox KH, Gonen M, Polotsky M, Hirsch BR, Schwartz LH. Response rate as a regulatory end point in single-arm studies of advanced solid tumors. JAMA Oncol 2016;2:772–9.

[123] Revicki DA, Osoba D, Fairclough D, Barofsky I, Berzon R, Leidy NK, et al. Recommendations on health-related quality of life research to support labeling and promotional claims in the United States. Qual Life Res 2000;9:887–900.

[124] Wilson MK, Collyar D, Chingos DT, Friedlander M, Ho TW, Karakasis K, et al. Outcomes and endpoints in cancer trials: Bridging the divide. Lancet Oncol 2015;16:e43–52.

[125] Chie W-C, Blazeby JM, Hsiao C-F, Chiu H-C, Poon RT, Mikoshiba N, et al. International cross-cultural field validation of an European Organization for Research and Treatment of

Cancer questionnaire module for patients with primary liver cancer, the European Organization for Research and Treatment of Cancer quality-of-life ques. Hepatology 2012;55:1122–9.

[126] Heffernan N, Cella D, Webster K, Odom L, Martone M, Passik S, et al. Measuring health-related quality of life in patients with hepatobiliary cancers: The functional assessment of Cancer Therapy-Hepatobiliary Questionnaire. J Clin Oncol 2002;20:2229–39.

[127] Yount S, Cella D, Webster K, Heffernan N, Chang CH, Odom L, et al. Assessment of patient-reported clinical outcome in pancreatic and other hepatobiliary cancers: The FACT Hepatobiliary Symptom Index. J Pain Symptom Manage 2002;24:32–44.

[128] Chie W-C, Blazeby JM, Hsiao C-F, Chiu H-C, Poon RT, Mikoshiba N, et al. Differences in health-related quality of life between European and Asian patients with hepatocellular carcinoma. Asia Pac J Clin Oncol 2017:e304–11.

[129] Llovet JM, Kudo M, Finn R, Galle PR, Blanc J, Okusaka T, et al. Ramucirumab As Second-Line Treatment in Patients with Hepatocellular Carcinoma (HCC) and Elevated Alpha-Fetoprotein (AFP) Following Sorafenib: Pooled Results from Two Global Phase 3 Studies (REACH-2 and REACH). Hepatology 2018;68, suppl.

[130] Park J-W, Kim YJ, Kim DY, Bae SH, Paik SW, Lee Y-J. Sorafenib with or without concurrent transarterial chemoembolization in patients with advanced hepatocellular carcinoma: a phase III STAH trial. J Hepatol 2018;In Press.

Figure legends

Figure 1: Median overall survival of treatment modalities assessed in phase III trials for advanced hepatocellular carcinoma.

Treatments with more than one dot represent all the results obtained from different clinical trials testing the same compound. Trials are colored based on whether the final result was positive for superiority (green), negative (red) or positive for non-inferiority (orange) for the primary endpoint (OS). Placebo appears in blue. Relevant inclusion/exclusion criteria that may impact on median OS are: no portal vein invasion [13], no pulmonary metastases [37], sorafenib tolerant [22], MET high [23] and AFP>400ng/ml [25].

Figure 2: Treatment strategy for advanced hepatocellular carcinoma. Adapted from Llovet et al. Nat Rev Clin Oncol 2018[6]**.**

Drugs in green have positive results from phase III trials with a superiority design (sorafenib in the first-line setting and regorafenib, cabozantinib and ramucirumab in the second-line setting). Drugs in orange have positive results from phase III trials with a non-inferiority design (lenvatinib in the first-line setting). Drugs in red have received accelerated approval from the FDA on the basis of promising efficacy results in phase II trials (nivolumab in the second-line setting). Key details of the patient populations are provided. AFP: Alpha-fetoprotein; BCLC: Barcelona Clinic Liver Cancer (classification); ECOG PS: Eastern Cooperative Oncology Group performance status; EHS: extrahepatic spread; HCV: hepatitis C virus; HR: hazard ratio; mRECIST: modified Response Evaluation Criteria In Solid Tumors; ORR: objective response rate; OS: overall survival.

Figure 3: Correlation between surrogate endpoints (PFS[A] and TTP[B]) and hard endpoint (OS).

Trial-level correlation between endpoints using criteria from the Institute for Quality and Efficiency in Health Care (IQWIG). R and R2 refers to the weighted Pearson coefficient between the HR of OS and the HR of the surrogate endpoint. IQWIG categorizes the strength of the correlation based on the value of R as low (R<0.7), moderate (R>0.7 to R<0.85) and high (R>0.85)[89]. Each dot represents one of the phase III clinical trials conducted on advanced HCC. Size of the dot is proportional to the total number of patients enrolled in the trial. Trials are colored based on whether the final result was positive, negative or non-inferior for the primary endpoint (OS). X and Y axis depict the value of the HR for the surrogate (TTP or PFS) and the hard endpoint (OS), respectively. Gray shaded areas represent the upper and lower limits of the 95% confidence intervals for the regression.

Figure 4: Correlation between surrogate endpoints PFS and TTP.

Trial-level correlation between endpoints. R and R2 refers to the weighted Pearson coefficient between the HR of PFS and the HR of TTP. Each dot represents one of the phase III clinical trials conducted on advanced HCC. Size of the dot is proportional to the total number of patients enrolled in the trial. Trials are colored based on whether the final result was positive, negative or non-inferior for the primary endpoint (OS). X and Y axis depict the value of the HR for the surrogate TTP and PFS, respectively. Gray shaded areas represent the upper and lower limits of the 95% confidence intervals for the regression.

Supplementary Figure 1: Correlation between objective response and OS.

Trial-level correlation between endpoints. R and R2 refers to the weighted Pearson coefficient between the HR of OS and the odds ratio of objective response. Each dot represents one of the phase III clinical trials conducted on advanced HCC. Size of the dot is proportional to the total number of patients enrolled in the trial. Trials are colored based on whether the final result was positive, negative or non-inferior for the primary endpoint (OS). X and Y axis depict the value of the odds ratio for objective response and the HR for OS, respectively. Gray shaded areas represent the upper and lower limits of the 95% confidence intervals for the regression.

Tables

Table 1: Phase III trials in advanced hepatocellular carcinoma conducted in the last decade.

Table 2: Studies analyzing associations between radiological response and survival in hepatocellular carcinoma.

Table 1

	Trial	Arms	N	ORR	TTP		PFS		OS	
					Median	HR	Median	HR	Median	HR
First-line	SHARP (7)	Sorafenib	299	2.3	5.5	0.58 (0.45 - 0.74)	NR		10.7	0.69 (0.55 - 0.87)
		Placebo	303	0.7	2.8				7.9	
	Asian-Pacific (8)	Sorafenib	150	3.3	2.8	0.57 (0.42 - 0.79)	NR		6.5	0.68 (0.50 - 0.93)
		Placebo	76	1.3	1.4				4.2	
	SUN1170 (9)	Sunitinib	530	6.6	4.1	1.13 (0.98 - 1.31)	3.6	1.13 (0.99 - 1.30)	7.9	1.30 (1.13 - 1.50)
		Sorafenib	544	6.1	3.8		3		10.2	
	BRISK-FL* (10)	Brivanib	577	12.0	4.2	1.01 (0.88 - 1.16)	NR		9.5	1.07 (0.94 - 1.23)
		Sorafenib	578	8.8	4.1				9.9	
	LIGHT (11)	Linifanib	514	10.1	5.4	0.76 (0.64 - 0.90)	4.2	0.81 (0.70 - 0.95)	9.1	1.05 (0.90 - 1.22)
		Sorafenib	521	6.1	4		2.9		9.8	
	SEARCH (12)	Sorafenib+Erlotinib	362	6.6	3.2	1.14 (0.94 - 1.37)	NR	1.11 (0.94 - 1.31)	9.5	0.93 (0.78 - 1.11)
		Sorafenib	358	3.9	4				8.5	
	REFLECT* (13)	Lenvatinib	478	24.1	8.9	0.63 (0.53 - 0.73)	7.4	0.66 (0.57 - 0.77)	13.6	0.92 (0.79 - 1.06)
		Sorafenib	476	9.2	3.7		3.7		12.3	
	SARAH (14)	Y90	237	15.2	NR		4.1	1.03 (0.85 - 1.25)	8	1.15 (0.94 - 1.41)
		Sorafenib	222	10.4			3.7		9.9	
	SIRveNIB (15)	Y90	182	16.5	6.1	0.88 (0.7 - 1.1)	5.8	0.89 (0.70 - 1.10)	8.8	1.10 (0.90 - 1.40)
		Sorafenib	178	1.7	5.4		5.1		10	
	EACH (16)	FOLFOX4	184	8.2	NR		2.93	0.62 (0.49 - 0.79)	6.4	0.80 (0.63 - 1.02)
		Doxorubicin	187	2.7			1.77		4.97	
	CALGB80802 (17)	Sorafenib+Doxorubicin	173	NR	NR		3.6	0.90 (0.72 - 1.20)	9.3	1.06 (0.80 - 1.40)
		Sorafenib	173	NR			3.2		10.5	
	SILIUS* (18)	Sorafenib+HAIC	103	36.3	5.3	0.65 (0.48 - 0.87)	4.8	0.75 (0.57 - 1.00)	11.8	1.01 (0.74 - 1.37)
		Sorafenib	103	17.5	3.5		3.5		11.5	
Second-line	BRISK-PS* (19)	Brivanib	263	9.9	4.2	0.56 (0.42 - 0.76)	NR		9.4	0.89 (0.69 - 1.15)
		Placebo	132	1.5	2.7				8.2	
	EVOLVE-1 (20)	Everolimus	362	2.2	3	0.93 (0.75 - 1.15)	NR		7.6	1.05 (0.86 - 1.27)
		Placebo	184	1.6	2.6				7.3	
	REACH (21)	Ramucirumab	283	7.1	3.5	0.59 (0.49 - 0.72)	2.8	0.63 (0.52 - 0.75)	9.2	0.87 (0.72 - 1.05)
		Placebo	282	0.7	2.6		2.1		7.6	
	RESORCE* (22)	Regorafenib	379	10.6	3.2	0.44 (0.36 - 0.55)	3.1	0.46 (0.37 - 0.56)	10.6	0.63 (0.50 - 0.79)
		Placebo	194	4.1	1.5		1.5		7.8	
	METIV-HCC (23)	Tivantinib	226	0.0	2.4	0.96 (0.74 - 1.25)	2.1	0.96 (0.75 - 1.22)	8.4	0.97 (0.75 - 1.25)
		Placebo	114	0.0	3		2		9.1	
	CELESTIAL (24)	Cabozantinib	470	3.8	5.4	0.41 (0.34 - 0.49)	5.2	0.44 (0.36 - 0.52)	10.2	0.76 (0.63 - 0.92)
		Placebo	237	0.4	1.9		1.9		8	
	REACH-2 (25)	Ramucirumab	197	4.6	3.02	0.43 (0.31 - 0.58)	2.8	0.45 (0.34 - 0.60)	8.5	0.71 (0.53 - 0.95)
		Placebo	95	1.1	1.61		1.6		7.3	
	ADI-PEG 20 (26)	ADI-PEG 20	424	NR	NR		2.6	1.18 (0.96 - 1.43)	7.8	1.02 (0.85 - 1.23)
		Placebo	211	NR			2.6		7.4	
	ReLive (27)	Doxorubicin Transdrug	263	0.8	NR		2.3	0.95 (0.74 - 1.22)	9.1	1.00 (0.78 - 1.28)
		Placebo	134	0.7			2.3		9	

*Radiological evaluation by mRECIST
NR = Non reported
Green = Positive for superiority ($p < 0.05$)
Orange = Positive for non-inferiority (upper 95% CI < 1.08)
Red = Negative for superiority or non-inferiority

Table 2

	Treatment	Study	N	Radiological response	Median OS (HR/OR OS) [R vs Non-R]	Prediction of OS
Local ablation	RFA	Cabibbo G et al (retrospective RFA)* (103)	151	CRR=78%	59.4m vs 26m (HR=0.3)	Yes (UV)
	RFA/PEI	Sala et al (retrospective)* (105)	282	CRR=68%	43m vs 28m (OR=0.58)	Yes (MV)
Chemo-embolization	TAE/TACE	Gillmore R et al (retrospective) (106)	83	ORR=57%	20.7m vs 13.3m (HR=0.58)	Yes (MV)
	TACE	Kim BK et al (retrospective) (107)	292	ORR=71.9%	33.8m vs 17.1m (HR=0.48)	Yes (MV)
	TACE	Jung ES et al (retrospective) (108)	114	ORR=63.3%	41.1m vs 20.7m (HR=0.31)	Yes (MV)
	DEB TACE	Prajapati HJ et al (retrospective) (109)	120	ORR=52.5%	28m vs 9.1m (HR=0.4)	Yes (MV)
Sorafenib	Sorafenib	Edeline J et al (retrospective) (98)	53	ORR=23%	18.2m vs 7.7m (NR)	Yes (UV)
	Sorafenib	Ronot M (retrospective)** (99)	82	ORR=28%	25.5m vs 5.7m (HR=0.19)	Yes (MV)
	Sorafenib	Takada J (retrospective) (100)	191	ORR=13.1%	≈21m vs ≈10m (NR)	Yes (UV)
Other systemic therapies	Brivanib	Lencioni R et al (phase III trial) (101)	226	ORR=11.5%	14.3m vs 9.4m (HR=0.48)	Yes (MV)
	Nintedanib	Meyer T et al (2 phase II trials) (102)	180	ORR=15.6%	16.7m vs 10.9m (HR=0.62)	Yes (MV)
	Nivolumab	El-Khoueiry A et al (phase II trial)*** (65)	145	ORR=14%	Non-reached vs 13.4m^ (NR)	Yes (UV)

CRR = Complete response rate
ORR = Objective response rate
R = Response (Objective or complete response)
HR = Hazard ratio
OR = Odds ratio
m = months
UV = Univariate
MV = Multivariate
NR = Non reported
*EASL criteria
**Non-R does not include stable disease
***RECIST 1.1 criteria
^0-25% reduction (17.7m); 0-25% increase (11.7m); ≥25% increase (8.9m).

Figure 1

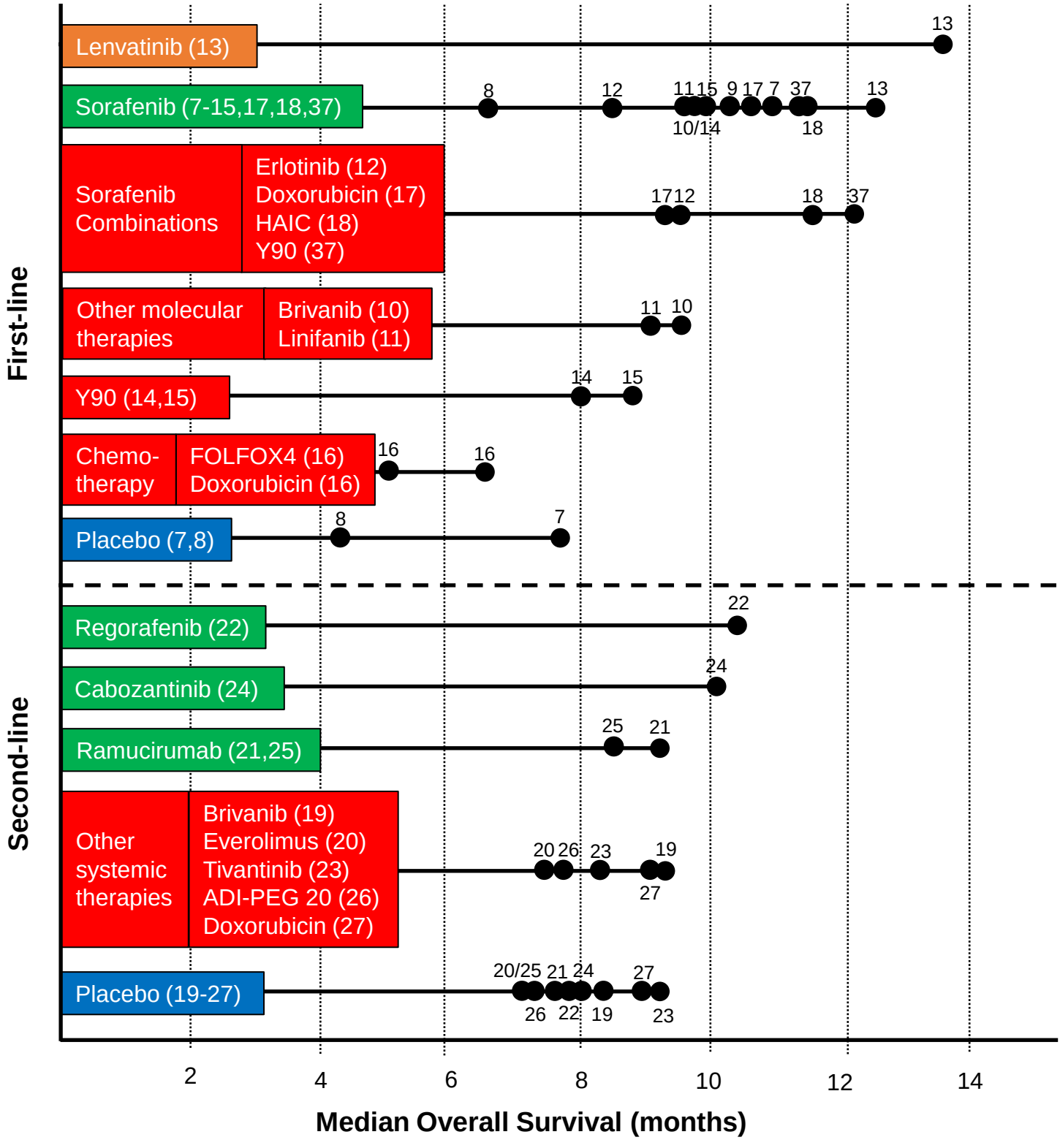

First-line
Second-line
Lenvatinib (13)
Sorafenib (7-15,17,18,37)
Sorafenib Combinations
Erlotinib (12)
Doxorubicin (17)
HAIC (18)
Y90 (37)
Other molecular therapies
Brivanib (10)
Linifanib (11)
Y90 (14,15)
Chemo-therapy
FOLFOX4 (16)
Doxorubicin (16)
Placebo (7,8)
Regorafenib (22)
Cabozantinib (24)
Ramucirumab (21,25)
Other systemic therapies
Brivanib (19)
Everolimus (20)
Tivantinib (23)
ADI-PEG 20 (26)
Doxorubicin (27)
Placebo (19-27)
2
4
6
8
10
12
14
Median Overall Survival (months)

Figure 2

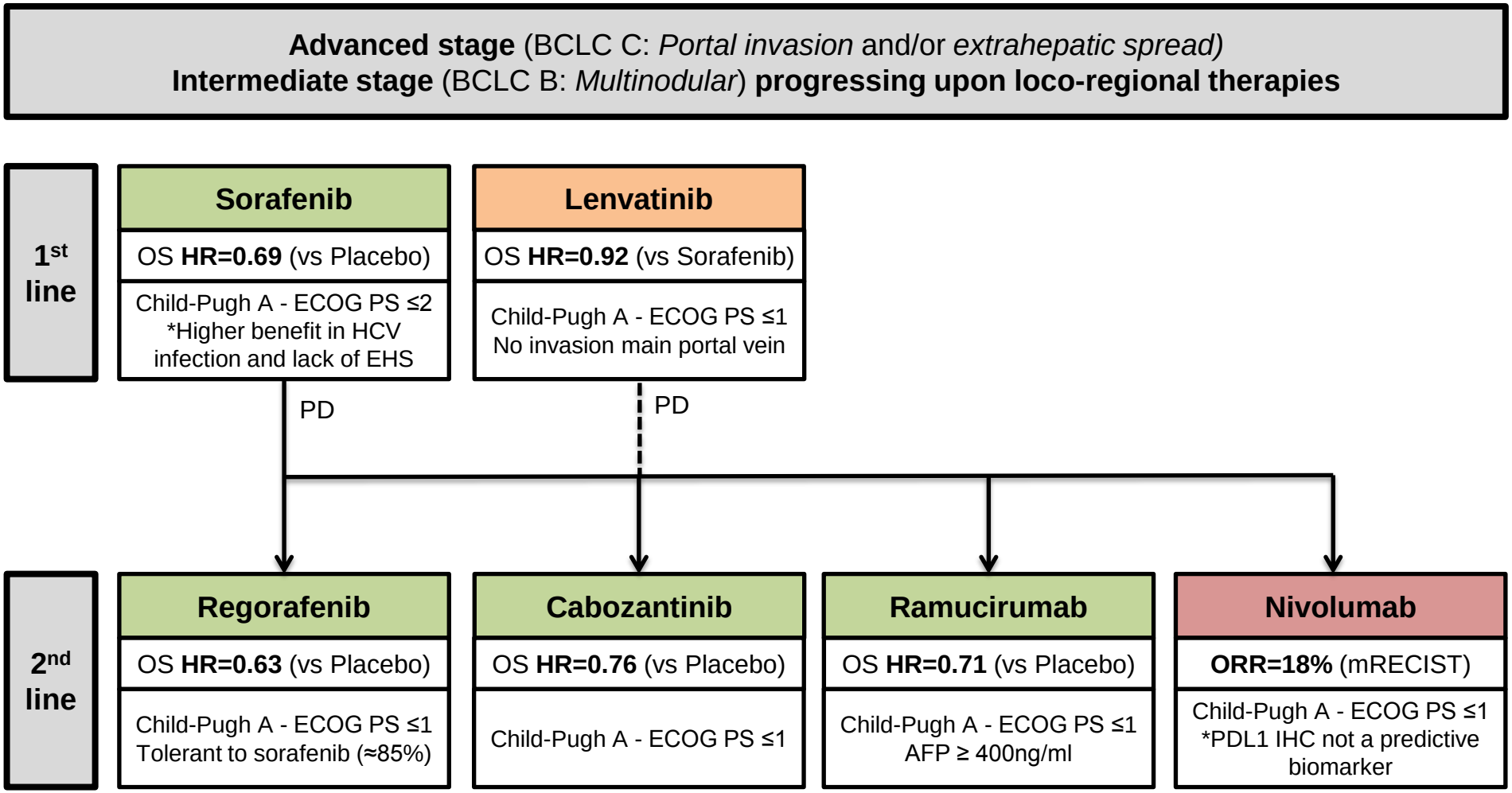

Advanced stage (BCLC C: Portal invasion and/or extrahepatic spread)
Intermediate stage (BCLC B: Multinodular) progressing upon loco-regional therapies
1st line
Sorafenib
OS HR=0.69 (vs Placebo)
Child-Pugh A - ECOG PS ≤2
*Higher benefit in HCV infection and lack of EHS
Lenvatinib
OS HR=0.92 (vs Sorafenib)
Child-Pugh A - ECOG PS ≤1
No invasion main portal vein
PD
PD
2nd line
Regorafenib
OS HR=0.63 (vs Placebo)
Child-Pugh A - ECOG PS ≤1
Tolerant to sorafenib (≈85%)
Cabozantinib
OS HR=0.76 (vs Placebo)
Child-Pugh A - ECOG PS ≤1
Ramucirumab
OS HR=0.71 (vs Placebo)
Child-Pugh A - ECOG PS ≤1
AFP ≥ 400ng/ml
Nivolumab
ORR=18% (mRECIST)
Child-Pugh A - ECOG PS ≤1
*PDL1 IHC not a predictive biomarker

Figure 3

Click here to download high resolution image

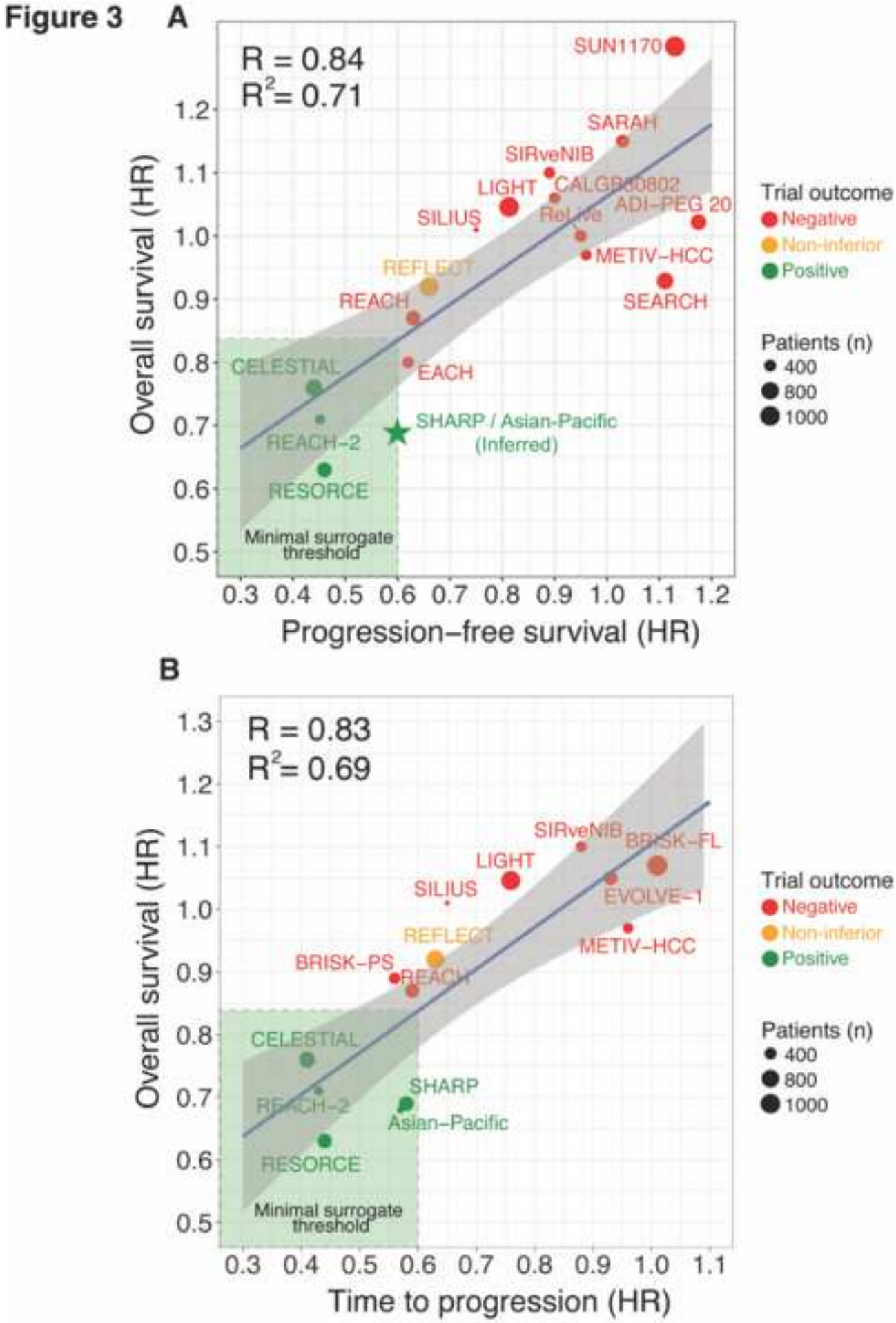
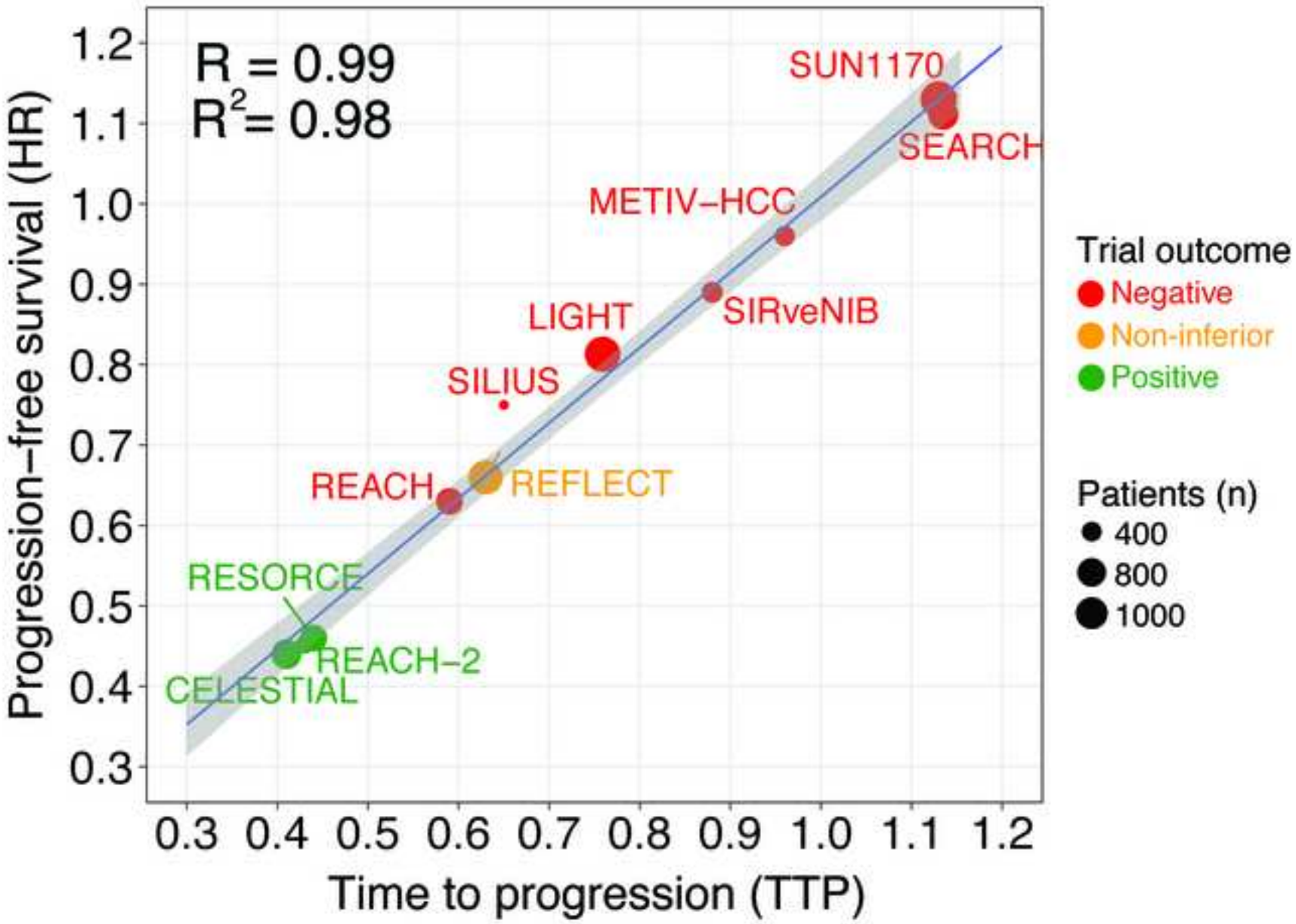


Figure 4

Click here to download high resolution image

Figure 4



Supplementary Figure 1

Click here to download Supplementary material: Supplementary Figure 1.tif

***ICMJE disclosure form**

Click here to download ICMJE disclosure form: COI_disclosure_RM_final.pdf

***ICMJE disclosure form**

Click here to download ICMJE disclosure form: COI_disclosure_AV_final.pdf

***ICMJE disclosure form**

Click here to download ICMJE disclosure form: COI_disclosure_JML_final.pdf