Expert Opinion

JOURNAL OF HEPATOLOGY

Endpoints in clinical trials for liver cancer and their value in evidence-based clinical decision making: An unresolved Gordian knot

Jordi Bruix*

Summary

The design and execution of clinical trials relies on strict definitions and criteria to avoid heterogeneous decisions by investigators at different sites. Ideally, definitions and decision making in clinical practice should mimic those implemented in trials, but this is not the case. Target populations are narrowly defined in trials, with the goal of evaluating activity and toxicity, and ultimately, demonstrating a survival benefit. In real-world practice, patients may not fit into the stringent inclusion/exclusion criteria of clinical trials. The evaluation of activity may also differ and the common policy to stop therapy upon progression may not be followed if progression is minor. Indeed, registration of progression may not reflect treatment failure or resistance. Parameters such as response according to RECIST criteria, time to progression and progression-free survival are not fully informative and cannot be assumed as a definitive surrogate for survival, which is the hardest endpoint in therapeutic cancer studies. This difference is because of the varying methods used to evaluate drug activity and tumour evolution, which ultimately dictates patient outcome. This expert opinion exposes the current discrepancies between research trials and clinical practice. Understanding the origin and limitations of such a conundrum should be the first step in refining the criteria that define drug activity, toxicity and treatment failure. Otherwise, evidence-based clinical practice and precision oncology will be an unattainable reality.

Keywords: Hepatocellular carcinoma; Cholangiocarcinoma; RECIST; Treatment; Precision oncology; Trial design; Evidence-based medicine.

Received 5 January 2021; received in revised form 21 January 2021; accepted 22 January 2021; available online 5 February 2021

Introduction

Clinical trials are the cornerstone of evidence-based medicine. Positive results of a trial with adequate design inform clinical practice guidelines, while negative results prevent the clinical application of medical interventions with unproven value. Several steps in trial design are key. The first is to identify a relevant problem to be investigated. This implies defining the ideal primary endpoint, setting the hypothesis and identifying a target population with well-defined inclusion/exclusion criteria. The second step includes: number of trial arms, use of placebo or active standard of care in the control arm, calculation of sample size according to predefined assumptions and optimal selection of stratification factors. The final step includes: follow-up/monitoring methodology, independent validation of events related to primary and secondary endpoints, and prospective definition of an adequate analysis plan.

These aspects are relevant in all phases of drug development, although primary endpoints vary. It is common to divide trials into separate phases: phase I trials concentrate on safety, phase II trials aim to identify activity signals that indicate the potential for efficacy, and phase III trials focus on survival benefit over the standard of care and provide the basis for regulatory approval by official agencies, which is based on an overall positive benefit/risk assessment. Several modifications to this approach have been proposed to speed up clinical development, but these are beyond the scope of this opinion piece. Obviously, endpoints should be clinically meaningful and not just a mere compendium of efficacy data lacking real usefulness for real clinical practice. An in-depth analysis of the criteria used in liver cancer trials highlights the major need to refine and improve current tools.[1] Indeed, the only follow-up event that has absolute precision is death. However, capturing survival data requires longer trial durations, increasing costs and delaying results. In addition, since some patients may receive post-trial sequential treatments, it has been suggested that long-term survival is no longer robust because of the beneficial or deleterious impact of such sequential therapies. This may be the case in cancer types with very long-term survival and plenty of therapeutic alternatives. However, survival of patients with advanced hepatocellular carcinoma (HCC) under systemic therapy is still dismal and limited agents are widely available. Therefore, using overall survival (OS) as a primary endpoint

BCLC Group, Liver Unit, Hospital Clínic, University of Barcelona, IDIBAPS, CIBEREHD, Spain

* Corresponding author. Address: BCLC Group, Liver Unit, Hospital Clínic, University of Barcelona, IDIBAPS, CIBEREHD, Villarroel 170, Barcelona, 08036, Spain.

E-mail address: jbruix@clinic.cat.

https://doi.org/10.1016/j.jhep.2021.01.033

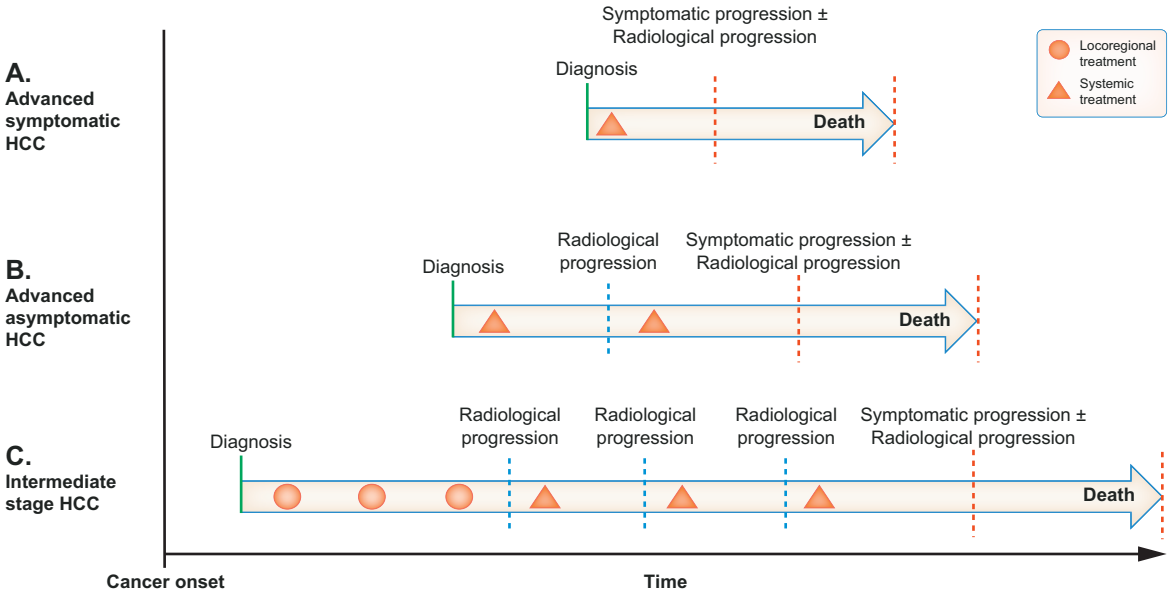

Fig. 1. Time to progression and PFS have increased because of earlier diagnosis and initiation of treatment well in advance of end-stage disease and death. (A) Cancer progression and death were very close in time when patients were diagnosed and treated at an advanced symptomatic stage. Cancer progression and thus PFS, could be valid surrogates for survival. (B) Earlier diagnosis at an asymptomatic advanced stage permits treatment initiation at an earlier evolutionary time point. Time to progression is the same as in A, but in such instances, survival after progression is increased, but merely because of earlier diagnosis (lead time bias). Thus, cancer progression irrespective of its pattern (and thus PFS), are less close to death and the surrogacy for survival is hampered. Survival is improved because of availability of treatments that confer survival benefit. (C) An even earlier diagnosis and treatment initiation exaggerate the distance between progression and death. The surrogate value of PFS for survival is no longer in place. Survival may further improve because of sequential administration of effective options. Note that additional confounders such as liver failure or severe treatment-related adverse events leading to death should also be considered. HCC, hepatocellular carcinoma; PFS, progression-free survival.

would not represent a significant increase in trial duration if balanced against the risk of a vulnerable result because of the use of a surrogate parameter for OS.[2]

In the next sections, I comment on the value and limitations of the criteria currently used in trials for either HCC or intrahepatic cholangiocarcinoma (iCC), while also describing limitations of their use in practice.

Adverse event registration

In oncology, adverse events (AEs) emerging during treatment are defined according to the Common Terminology Criteria for Adverse Events (CTCAE).[3] AEs are stratified as treatment related or unrelated, while severity is graded into 5 grades, with grade 5 reflecting death. In studies including patients with liver cancer in whom underlying liver disease is common, the evaluation of AEs related to liver function is challenging.[4] Liver disease may progress irrespective of tumour evolution and infiltrative tumour progression may impair liver function without evident tumour growth on imaging. Indeed, the definitions used to assess drug-induced liver damage differ from those used in the Hepatology realm.[4,5] At the same time, novel therapies such as immunotherapy whose action is based on non-cytotoxic mechanisms may elicit AEs that may not be properly captured, because terminology does not match the recommendations established by each medical speciality (Hepatology, Cardiology, Endocrinology, Ophtalmology, *etc.*) for the investigation and management of organ-specific alterations.[6] Finally, AEs are usually presented as a total number without describing their occurrence over time. Thus, treatments that improve survival may appear to cause a disproportionate number of AEs, as data accumulate over a longer follow-up time than for alternative treatments that confer less survival benefit. Phase I trials usually have a limited sample size that allows for close analysis of AEs and enables study investigators to make clear conclusions on safety. By contrast, phase II-III trials have larger sample sizes and the role of Data Safety Monitoring Committees is key to carefully inspect the trial data to ensure the safety of the tested intervention. AEs identified within trials are reported in detail, but when agents reach practice it is necessary to consider both expected events as well any event that could be treatment related. This is the role of pharmacovigilance organisations after drug approval. Again, grading of AEs in practice does not follow the CTCAE and it may take years to establish causality.

AEs and their severity will likely impact on quality of life. This may be captured by patient-reported outcomes (PROs) that enable the quantification of the perceived deterioration or preservation in a patient's quality of life.[7] PROs have been proposed as a way of comparing the value of care offered by different teams, although their use in

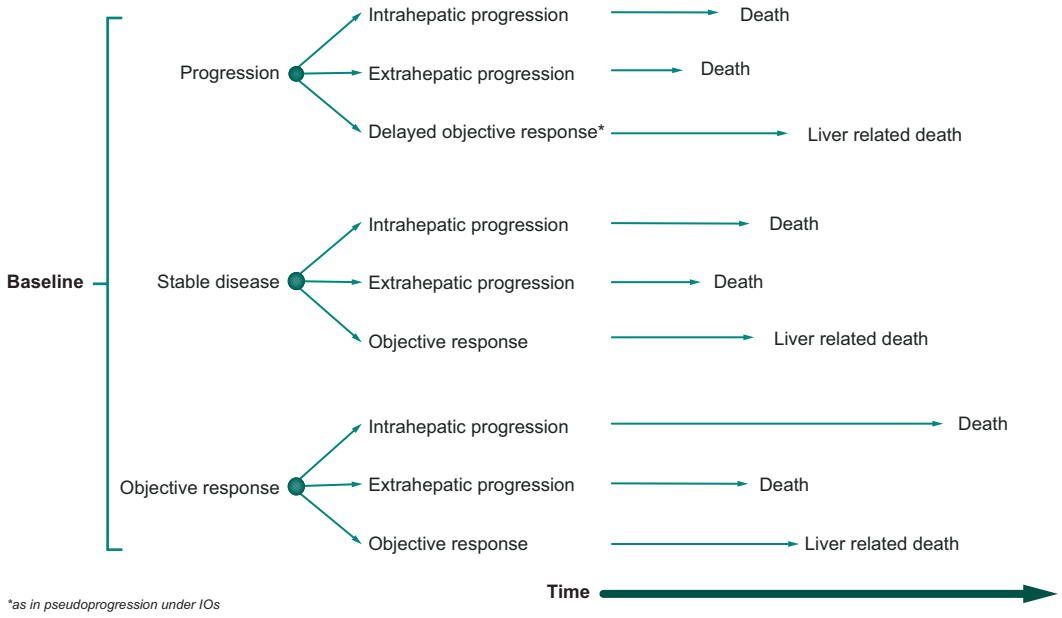

Fig. 2. The faulty surrogate value for survival of tumour progression at imaging. Tumour evolution may be highly heterogeneous. If PFS is the endpoint, the detection of progression defines the time of this combined event. However, progression may have distinct patterns with significantly different impact on post progression survival. Thus, survival after registering progression may be very heterogeneous as represented by arrows of different length. In addition, some patients initially registered as having progressive or stable disease may evolve to objective response (*as seen under immunotherapy), while other patients may die without tumour recurrence/progression due to liver-related complications (variceal bleeding or infections leading to acute-on-chronic liver failure). As a result, the value of PFS as a surrogate for overall survival is not robust. PFS, progression-free survival.

conventional practice is not common. One issue is that the comparison of PROs across populations is hampered by economic and cultural heterogeneity between communities.

Assessment of activity and potential signals in phase II trials

The concept that underlies the evaluation of treatment efficacy stems from the idea that elimination of cancer or reduction of tumour burden is a positive event that ultimately should translate into survival benefit. Accordingly, assessment of treatment is based on evolutionary changes in radiological tumour burden. Reduction is response, no change is stable disease, and growth beyond a given cut-off is progression. In the early days of cancer treatment this stratification was done by palpation of abdominal masses or lymph nodes. This prompted an interobserver agreement study using rubber balls that mimicked abdominal masses or lymph nodes of different sizes.[8] Based on its results a decrease in size >50% would be classified as response while progression would be registered when the increase would exceed 25%.[8] These values became the cut-offs for the seminal WHO criteria[9] that were later replaced by the Response Evaluation Criteria in Solid Tumours (RECIST) criteria.[10] Cut-offs were modified to accommodate the change from area to diameter (30% decrease from baseline for response and 20% increase from nadir for progression).[10] In addition to the evaluation of target tumour lesions, these definitions incorporated the concept of "*best response*" achieved and of "*duration of response*". The emergence of new tumour sites in any location is registered as progression, which led to the incorporation of "time to progression (TTP)". "Depth of response" may provide more insight about activity.[1] However, a major depth of response followed by early recurrence and progression to death refutes the value of depth as a fully informative parameter. This is exemplified by surgery in advanced cancer: response may be complete, but recurrence may follow fast.

It makes sense that tumour growth and dissemination predict poor prognosis as this is what leads to death. Decades ago, patients were diagnosed at an advanced symptomatic stage and progression would translate into early death (Fig. 1). Currently, diagnosis is done at earlier asymptomatic stages and even if progression occurs, death may occur far later. Therefore, a robust correlation between progression and death may be missing. Indeed, the pattern of progression (Fig. 2) may be more relevant to predict outcome as shown by Reig *et al.*[11] and extensively validated thereafter.[12,13] This should not be unexpected as the impact on prognosis of a tiny new hepatic nodule is obviously different from that of distant metastasis.

Thus, what is the value of response? Reduction in tumour burden is a marker of activity, but activity is not to be ruled out in the absence of response as

Table 1. Relevant differences between research trials and clinical practice in liver cancer.

	Clinical trial	Clinical practice
Tumour burden evaluation	Regular intervals	Follow-up may not be strict and regular
Tumour assessment	Assessed by RECIST	May not be assessed by RECIST
Pattern of progression	Up to now not considered	Heterogeneously considered by physicians
Treatment interruption	At radiological progression	May be maintained beyond progression
Adverse events and toxicity evaluation	Graded by CTCAE	Graded according to clinical terminology not always concordant with CTCAE
Patient reported outcomes	Collected during trial duration	Usually not collected during clinical visits

CTCAE, Common Terminology Criteria for Adverse Events; RECIST, Response Evaluation Criteria in Solid Tumours.

exemplified by sorafenib.[14] It induced a marginal number of responses, but it delayed tumour progression and improved survival compared to placebo. The same has been observed with almost all agents with proven survival benefit in HCC. It is known that locoregional treatments such as ablation, chemoembolisation or radiation, can produce massive tumour necrosis without changing lesion size. RECIST would register stability rather than response, but a major or complete pathological response is often seen after such locoregional modalities.[15] This was the basis for the EASL response criteria for HCC that take into account the extent of tumour necrosis, recognised by the absence of contrast enhancement at dynamic imaging, as a signal of activity.[16] mRECIST criteria used this validated concept for locoregional therapies and it was hypothesised to be informative for systemic agents.[17] However, agents with antiangiogenic activity (tyrosine kinase inhibitors and anti-vascular endothelial growth factor antibodies) induce arterial vasoconstriction with decreased mesenteric blood flow and reduced arterial inflow into the tumour.[18] This decreases arterial contrast uptake at dynamic imaging, which resembles tumour necrosis, and can mislead physicians to infer an antitumour activity. Reducing the vascularisation of a tumour is not a response and such a flaw leads to the overestimation of antitumoural activity in early phase trials. Of course, some patients may present huge tumour necrosis, but this is not the rule.

Registration of stable disease or progression under immunotherapy is also not fully meaningful. Immunotherapy may induce “pseudoprogression” that is not confirmed later or is even followed by a response. At the same time, dissociate response patterns are sometimes observed, wherein some tumour sites may present growth while others may shrink.[19] All these observations have led to the development of the immune-RECIST system for response evaluation but it still needs further refinement.[20]

Detection by imaging of a new tumour site in any location after initial response is registered as progression and this is assumed to reflect treatment failure or resistance. In most research trials it triggers treatment interruption. This is highly controversial as the heterogeneous tumour clones may result in some of them being under control, while others progress because they are not affected by treatment. The WHO and RECIST models acknowledge that detection of progression should not lead to treatment interruption, which should instead be based on clinical judgement.[21,22] Thus, TTP is not thought to be an informative parameter. Indeed, in liver cancer, the definition of progression due to new hepatic nodules risks false positives because of the detection of macroregenerative nodules within a cirrhotic liver. Such nodules are not highly vascularised. The associated risk of false positives led to the creation of the radiology charter for the pivotal sorafenib trial.[23] Therein, it was stated that a new nodule should exceed 10 mm in diameter and show arterial vascularisation or progressive growth to be classified as malignant. mRECIST proposed a more stringent definition by requiring arterial enhancement followed by washout in the venous phases,[17] thus mimicking the diagnostic criteria for the primary nodule. This demand is not implemented in clinical practice after resection, ablation, locoregional interventions or even, systemic therapy. Furthermore, prospective studies have shown that using mRECIST does not increase TTP: it delays confirmation, but date of progression is the one of first detection.[12]

It is striking that the definition of patient status and treatment decisions upon the registration of progression differ sharply between research trials and clinical practice (Table 1). In most research trials progression triggers treatment interruption, but at the same time they allow exceptions to be made if the investigators feel that the patient would still benefit from treatment. Similarly, in clinical practice the term “marginal progression” has emerged and has been used to maintain treatment.[22] This discrepancy between research and practice is impossible to reconcile. If progression is taken as treatment failure and thus, the basis for treatment interruption, this should be applied in real life to avoid exposure of patients to agents that are said to have lost efficacy and may still induce AEs and impair quality of life. However, if tumour progression of small intensity should not be considered as a treatment failure, as often occurs in clinical practice, then trials should add such consideration in their design. This perspective is relevant for physicians involved in trials, as it may become an ethical issue to make contradictory treatment decisions depending on the setting.

Predictors of outcome and surrogates for survival

It is important to distinguish between baseline parameters that may predict a better or worse survival, and the evolutionary events that may become surrogate markers for different survival outcomes. Evolutionary events that emerge after treatment initiation (adverse events, increase of tumour markers, tumour progression) should not be included as baseline data to prevent flaws. Multivariate analysis with time-dependent models and landmark analysis for events that appear at a defined time point are valid tools to avoid such flaws. Time-dependent multivariate analysis should include both the radiological evolution in tumour burden (response, stability, progression), as well as liver function parameters and physical status at the same time point. Otherwise, the sole parameter explored is limited to cancer evolution, while others may be more powerful drivers of survival or be the sole relevant parameter. Nonetheless, the critical point in survival prediction according to tumour evolution is that the definitions employed for response may not have the required granularity to identify activity (this is why study reports frequently incorporate spider or waterfall graphs) and that progression should be stratified by its pattern. It is also important to separately assess response, stabilisation or progression, and not lump together response and stabilisation, and compare them to progression. Lumping together stabilisation and progression and comparing them to response should also be avoided. These do not explore if the survival benefit differs between each evolutionary possibility.

There are strict rules to establish if an event is a surrogate for a later event.[24,25] The association between surrogates (response to treatment, TTP or progression-free survival [PFS]) and the effect of the treatment on the true endpoint (survival) must be shown at the individual and at the trial level. In such studies, it does not suffice to have a statistically significant correlation "r" with 95% CI; instead, a coefficient of determination (r^2) of proper strength should be used, with the correlation strength determined by the lower 95% CI value. Also, when available, the surrogate threshold might be useful to identify the minimal effect on the surrogate endpoint which might guarantee a benefit on the true endpoint.[24,25] Unfortunately, up to now, there are no robust data to support response rate, TTP or PFS as surrogates for survival. All suggestions await proper validation with adequate statistics and interpretation. Until such evidence becomes available, any trial designed with PFS as the primary endpoint should be seen as vulnerable and follow-up survival should be demanded. Official agencies may grant conditional approval, but at the same time demand survival results.

In summary, definitions within clinical trials are important for homogenous patient management, as they prevent individual physicians at different sites from making divergent decisions that risk generating heterogeneity. However, the interpretation of trial findings and their impact on relevant clinical decisions (drug activity, drug interruption, information about change in prognosis and need for alternative approaches) should be based on the scientific background and be equally valid and applied in clinical practice. Everyone involved in trial design and analysis should stay on the side of scientific robustness. If this is not the case, the core concept of evidence-based medicine and precision oncology will be at risk.

Abbreviations

AEs, adverse events; CTCAE, Common Terminology Criteria for Adverse Events; HCC, hepatocellular carcinoma; mRECIST, modified RECIST; OS, overall survival; PFS, progression-free survival; PROs, patient-reported outcomes; RECIST, Response Evaluation Criteria in Solid Tumours; TTP, time to progression.

Financial support

JB received grant support from Instituto de Salud Carlos III (PI18/00768), AECC (PI044031), and WCR (AICR) 16-0026. CIBERehd is funded by the Instituto de Salud Carlos III.

Conflict of interest

Consultancy: AbbVie, Adaptimmune, Arqule, AstraMedimmune, Basilea, Bayer, Bio-Alliance, BMS, BTG- Biocompatibles, Eisai, Gilead, Incyte, Ipsen, Kowa, Lilly, MSD, Nerviano, Novartis, Polaris, Quirem, Roche, Sirtex, Sanofi, Terumo. Research grants (Institutional): Bayer, BTG. Educational grants: Bayer, BTG. Paid conferences: Bayer, BTG, Ipsen and Roche. Paid talks: Bayer-Shering Pharma, BTG, Eisai, Terumo, Sirtex, Ipsen.

Please refer to the accompanying ICMJE disclosure forms for further details.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhep.2021.01.033.

References

Author names in bold designate shared co-first authorship

[1] Bruix J, da Fonseca LG, Reig M. Insights into the success and failure of systemic therapy for hepatocellular carcinoma. Nat Rev Gastroenterol Hepatol 2019;16:617–630. https://doi.org/10.1038/s41575-019-0179-x.

[2] Chen EY, Joshi SK, Tran A, Prasad V. Estimation of study time reduction using surrogate end points rather than overall survival in oncology clinical trials. JAMA Intern Med 2019;179:642–647. https://doi.org/10.1001/jamainternmed.2018.8351.

[3] US Department of Health and Human Services, National Institutes of Health NCI. Common terminology criteria for adverse events: (CTCAE) Version 5. 0 2017. https://ctep.cancer.gov/protocoldevelopment/electronic_applications/docs/CTC AE_v5_Quick_Reference_8.5x11.pdf. [Accessed 22 December 2020].

[4] Sherman M. Assessing toxicity in drug trials in liver disease. Semin Liver Dis 2021. https://doi.org/10.1055/s-0041-1723030 (In press).

[5] Palmer M, Regev A, Lindor K, Avigan MI, Dimick-Santos L, Treem W, et al. Consensus guidelines: best practices for detection, assessment and management of suspected acute drug induced liver injury occurring during clinical trials in adults with chronic cholestatic liver disease. Aliment Pharmacol Ther 2020;51:90–109. https://doi.org/10.1111/apt.15579.
[6] **Londoño MC**, **Reig M**. Multidisciplinary clinical approach to cancer patients with immune-related adverse events induced by checkpoint inhibitors. Cancers (Basel) 2020;12:1–22. https://doi.org/10.3390/cancers12113446.
[7] Coens C, Pe M, Dueck AC, Sloan J, Basch E, Calvert M, et al. International standards for the analysis of quality-of-life and patient-reported outcome endpoints in cancer randomised controlled trials: recommendations of the SISAQOL Consortium. Lancet Oncol 2020;21:e83–96. https://doi.org/10.1016/S1470-2045(19)30790-9.
[8] Moertel CG, Hanley JA. The effect of measuring error on the results of therapeutic trials in advanced cancer. Cancer 1976;38:388–394. https://doi.org/10.1002/1097-0142(197607)38:1<388::AIDCNCR2820380156>3.0.CO;2-A.
[9] Miller AB, Hoogstraten B, Staquet M, Winkler A. Reporting results of cancer treatment. Cancer 1981;47:207–214. https://doi.org/10.1002/1097-0142(19810101)47:1<207::AID-CNCR2820470134>3.0.CO;2-6.
[10] 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;45:228–247. https://doi.org/10.1016/j.ejca.2008.10.026.
[11] 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;58:2023–2031. https://doi.org/10.1002/hep.26586.
[12] Bruix J, Qin S, Merle P, Granito A, Huang YH, 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. https://doi.org/10.1016/S0140-6736(16)32453-9.
[13] Rimassa L, Assenat E, Peck-Radosavljevic M, Pracht M, Zagonel V, Mathurin P, et al. Tivantinib for second-line treatment of MET-high, advanced hepatocellular carcinoma (METIV-HCC): a final analysis of a phase 3, randomised, placebo-controlled study. Lancet Oncol 2018;19:682–693. https://doi.org/10.1016/S1470-2045(18)30146-3.
[14] Llovet JM, Ricci S, Mazzaferro V, Hilgard P, Gane E, Blanc J-F, et al. Sorafenib in advanced hepatocellular carcinoma. N Engl J Med 2008;359:378–390. https://doi.org/10.1056/nejmoa0708857.
[15] Forner A, Ayuso C, Varela M, Rimola J, Hessheimer AJ, De Lope CR, et al. Evaluation of tumour response after locoregional therapies in hepatocellular carcinoma: are response evaluation criteria in solid tumours reliable? Cancer 2009;115:616–623. https://doi.org/10.1002/cncr.24050.
[16] Bruix J, Sherman M, Llovet JM, Beaugrand M, Lencioni R, Burroughs AK, et al. Clinical management of hepatocellular carcinoma. Conclusions of the Barcelona-2000 EASL conference. J Hepatol 2001;35:421–430. https://doi.org/10.1016/S0168-8278(01)00130-1.
[17] Lencioni R, Llovet JM. Modified recist (mRECIST) assessment for hepatocellular carcinoma. Semin Liver Dis 2010;30:52–60. https://doi.org/10.1055/s-0030-1247132.
[18] Bruix J, Reig M, Sangro B. Assessment of treatment efficacy in hepatocellular carcinoma: response rate, delay in progression or none of them. J Hepatol 2017;66:1114–1117. https://doi.org/10.1016/j.jhep.2017.02.032.
[19] Rimola J, Da Fonseca LG, Sapena V, Perelló C, Guerrero A, Simó MT, et al. Radiological response to nivolumab in patients with hepatocellular carcinoma: a multicenter analysis of real-life practice. Eur J Radiol 2020:109484. https://doi.org/10.1016/j.ejrad.2020.109484.
[20] 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–e152. https://doi.org/10.1016/S1470-2045(17)30074-8.
[21] Litière S, Collette S, De Vries EGE, Seymour L, Bogaerts J. RECIST-learning from the past to build the future. Nat Rev Clin Oncol 2017;14:187–192. https://doi.org/10.1038/nrclinonc.2016.195.
[22] Oxnard GR, Morris MJ, Hodi FS, Baker LH, Kris MG, Venook AP, et al. When progressive disease does not mean treatment failure: reconsidering the criteria for progression. J Natl Cancer Inst 2012;104:1534–1541. https://doi.org/10.1093/jnci/djs353.
[23] Reig M, Darnell A, Forner A, Rimola J, Ayuso C, Bruix J. Systemic therapy for hepatocellular carcinoma: the issue of treatment stage migration and registration of progression using the BCLC-refined RECIST. Semin Liver Dis 2014;34:444–455. https://doi.org/10.1055/s-0034-1394143.
[24] Buyse M, Sargent DJ, Grothey A, Matheson A, De Gramont A. Biomarkers and surrogate end points - the challenge of statistical validation. Nat Rev Clin Oncol 2010;7:309–317. https://doi.org/10.1038/nrclinonc.2010.43.
[25] Demets DL, Psaty BM, Fleming TR. When can intermediate outcomes be used as surrogate outcomes? JAMA - J Am Med Assoc 2020;323:1184–1185. https://doi.org/10.1001/jama.2020.1176.