Methodological assessment of HCC literature

G. Daniele[1*], N. Costa[2], V. Lorusso[3], J. Costa-Maia[4], I. Pache[5] & M. Pirisi[6]

[1]Clinical Trials Unit, National Cancer Institute of Naples, Naples; [2]University Hospital and Clinic, Coimbra, Portugal; [3]National Cancer Research Centre, Istituto Tumori Giovanni Paolo II, Bari, Italy; [4]Department of Surgery, S. Joao Medical Center, Porto, Portugal; [5]Division of Gastroenterology and Hepatology, Centre Hospitalier Universitaire Vaudois, University of Lausanne, Lausanne, Switzerland; [6]Department of Translational Medicine, Università del Piemonte Orientale 'Amedeo Avogadro', Novara, Italy

Despite the fact that the hepatocellular carcinoma (HCC) represents a major health problem, very few interventions are available for this disease, and only sorafenib is approved for the treatment of advanced disease. Of note, only very few interventions have been thoroughly evaluated over time for HCC patients compared with several hundreds in other, equally highly lethal, tumours. Additionally, clinical trials in HCC have often been questioned for poor design and methodological issues. As a consequence, a gap between what is measured in clinical trials and what clinicians have to face in daily practice often occurs. As a result of this scenario, even the most recent guidelines for treatment of HCC patients use low strength evidence to make recommendations. In this review, we will discuss some of the potential methodological issues hindering a rational development of new treatments for HCC patients.

Key words: hepatocellular carcinoma, methodology, observational studies, randomized, clinical trials, sorafenib

introduction

Hepatocellular carcinoma (HCC) is the sixth most frequent tumour and represents the third leading cause of cancer death worldwide [1]. Differences in HCC risk factors and aetiology exist among different world regions (i.e. more related to hepatitis B virus (HBV) in Asian/African countries and to hepatitis C virus (HCV) and non-infectious in Western countries) and the incidence of this tumour is increasing mostly in Western countries along with the spread of HCV infection. Despite the fact that the HCC represents a major health problem, very few interventions are available for this disease. In fact, while potentially curative treatment options are available to localised disease patients, only one drug is approved for the treatment of advanced disease. In particular, the multikinase inhibitor sorafenib has been recently approved as first-line treatment for HCC patients, based on the results of two separate phase III trials conducted in Western and Asian countries, respectively [2, 3]. However, even more dismal is the fact that very few interventions have been thoroughly evaluated over time for HCC patients compared with several hundreds in other, equally highly lethal, tumours [4]. Only 16 phase III trials evaluating systemic treatments for advanced disease are active by June 2012 (from clinicaltrials.gov—last visit 15th June), with almost 50% of these evaluating sorafenib in special populations [there is only one study in Child-Pugh (CP) B patients] or combined with local treatment. Of the remaining seven trials three are not restricted to HCC patients. Additionally, clinical trials have often been questioned for poor design and methodological issues. As a consequence,

a gap between what is measured in clinical trials and what clinicians have to face in daily practice often occurs. As a result of this scenario, even the most recent guidelines for treatment of HCC patients use low strength evidence to make recommendations [5].

A seminal paper recently highlighted several factors potentially contributing to this scenario and tried to establish the bases of a common framework to improve the clinical investigation design in HCC [6]. In this review, we will discuss some of the potential methodological issues hindering a rational development of new treatments for HCC patients.

clinical development of anticancer drugs

Rational clinical development of new treatments for a particular tumour follows subsequent phases of investigation culminating with the regulatory approval. Phase I studies are usually small studies aimed to find the dose and the schedule for further investigations and to initially evaluate safety. The optimal design of phase I trials is aimed to minimize the number of patients treated with low, sub-therapeutic doses of the drug and those exposed to high, potentially toxic doses. Phase II trials are primarily aimed at the evaluation of the activity, tolerability and safety of the experimental drug in a larger population than in phase I. Other objectives of phase II trials include the description of the tolerability and the safety of the drug in a larger population than phase I. Multiple phase II trials in different tumour populations provide the direction of the further development by screening the tumour type in which the design of phase III trials is worthwhile. Randomised phase II trials are important tools to inform the go/non-go decision through phase III trials. However, they are far from being a definitive comparison between two treatments, nor are powered to support it. Phase III trials represent the most

---

Correspondence to: Dr G. Daniele, Clinical Trials Unit, National Cancer Institute of Naples, Via M. Semmola 80131, Naples, Italy. Tel: +39-0815903448; Fax: +39-0817732938; E-mail: gennaro.daniele@usc-intnapoli.net

robust and definitive test of the efficacy (i.e. how the patient will benefit from the treatment) of a drug in a particular tumour type. The evaluation of their results in terms of applicability to the population of patients encountered in clinical practice and of the extent of benefit the patients can derive from the treatment strongly relies upon variables such as study population, end points and study design (internal and external validity).

The successful progression of a new treatment to the approval is conditioned by the appropriate and informative conduction of all the phases. Study population, end points and study design acquire several peculiarities in HCC. Discussion about these points will be the main topic of the following sections.

selection of study population

Rational development of new anticancer targeted therapeutics moves through testing biological hypotheses, to find the drug levels needed to effectively inhibit the target (proof of principle) and to demonstrate a meaningful activity in appropriately selected population (proof of concept) [7]. Recent findings have emphasised the importance of selecting patients to treat with a particular drug in order to maximise the benefit they can derive from. Epidermal growth factor receptor (EGFR) mutations in predicting sensitivity to tyrosine kinase inhibitors [8], BRAF V600 mutation with vemurafenib in melanoma [9] or BRCA1/2 mutations with PARP inhibitors in ovarian cancer [10] are only few examples of this story started with imatinib and trastuzumab almost three decades ago.

Early-phase clinical trials represent the moment in which most of these biological hypotheses are tested [11]. Selection of patients based on clinical or molecular features in these trials might foster further research in a particular field and test whether effective target inhibition is achieved at therapeutic doses as well as address potential pharmacokinetic or pharmacodynamic variability among patients. Recent data have demonstrated the feasibility of this approach [12]. In the later stage of clinical investigation, the appropriate selection of the study population is fundamental to warrant the transferability of study results to the general population of patients (external validity). Several factors contribute in pursuing the rational selection of the HCC population for clinical trials and these factors are often critical points in published trials.

With these premises, the factors to be taken into account in selecting a study population in HCC include the availability of tissue samples, the prognosis of patients with HCC and the liver function. Ethnicity and other factors will be also briefly discussed.

tissue availability and biomarkers developing

Most predictive biomarkers are tissue-derived; multiple, sequential biopsies are often required for biomarker evaluation and validation. 'Non-invasive' diagnostic criteria for HCC have been established in 2005 [13]. Mainly due to the risk of seeding during the biopsy, quantified by Stigliano et al. [14],

and to the overall poor level of consensus among pathologists in the evaluation of the tumour grade, histological diagnosis has been largely replaced by radiological features [i.e. the contrast uptake during the arterial phase and rapid washout during the venous or the delayed phase at computed tomography (CT) or magnetic resonance imaging (MRI)]. As a result of this recommendation, very few patients have a tissue sample at diagnosis and very few studies (only three in the Physician Data Query database by June 2012; see http://www.cancer.gov/cancertopics/pdq/cancerdatabase) involving HCC patients required histological confirmation of the tumour as inclusion criteria. This paucity of tissue samples for most HCC patients is clearly a major issue for the rational development of new targeted agents. In fact, albeit the evaluation of potential predictive markers of response to particular targeted agents (i.e. drug target) can be warranted even with the few patients who have the tissue samples, their validation requires larger cohorts that are absolutely unrealistic in this scenario [15].

prognostic evaluation in selecting patients

Selecting patients with the same prognostic outcome is also critical to the rational clinical trials design. Generally, the prognostic evaluation of cancer patients includes both disease extent assessment and other relevant prognostic variables (i.e. liver function) [16]. Liver function will be extensively discussed below. Several staging systems have been proposed. The Barcelona Clinic Liver Cancer (BCLC) system represents an a priori stratification of HCC patients into four main categories according to the disease extent, liver function, performance status (PS), portal pressure and co-morbidities. Although its prospective external validation is lacking and it includes some subjective and complex investigations (PS and portal pressure, respectively), BCLC flow provides a treatment pathway for patients included in each category [5, 17]. This unique feature made it the most extensively used staging system both in clinical practice and in clinical trials. Other systems have been developed [16]. Among these, the Cancer of the Liver Italian Programme (CLIP) [18, 19] is widely used worldwide. In particular, the CLIP score has been retrospectively developed in a large population of HCC patients [18] and prospectively validated in an external population of HCC patients [19]. Notwithstanding some limitations (i.e. highest values in populations where the disease burden at diagnosis is low thanks to early detection), the CLIP score demonstrated superiority in prognostic assessment when compared with other systems like Okuda [19]. However, the validity of each score against the others is still under debate and as previously highlighted, with some exceptions, seems strongly dependent upon the population considered [20]. In this perspective, ethnicity is another important factor to be taken into account and it will be discussed later.

liver function

The vast majority of the HCC patients have underlying liver disease that can ultimately affect the liver function

independently from the HCC. Liver function deserves a particular consideration in designing clinical trials for patients with HCC for at least two reasons. First, the prognosis of severely impaired cirrhotic patients strongly relies upon the liver function irrespective of HCC [21, 22]. Therefore, given the worsening prognosis of the transition between compensated and decompensated cirrhosis, the relative benefit of a treatment declines as much as the liver function loss determines the death of the patient rather than the tumour progression [23]. Second, an impaired liver function alters, though at variable degrees, the drug tolerability and activity, potentially due to an altered metabolism of an active compound (including altered activation of prodrugs). At the mention of the latter point, a recent phase I trial of sorafenib in a population of liver and renal function impaired cancer patients (including 17 HCC patients) found that the recommended dose of sorafenib decreases with organ function worsening [24]. In particular, for patients with moderate [bilirubin >1.5 × ULN to ≤3 × ULN, any aspartate aminotransferase (AST), albumin normal] to severe (bilirubin >3 × ULN to 10 × ULN, any AST, albumin normal) up to very severe (albumin ≤2.5 mg/dl, any bilirubin, any AST) hepatic dysfunction, the recommended dose ranged from 200 mg twice a day to <200 mg every third day up to 200 mg once a day, respectively. Interestingly, the trials did not find any significant correlation between sorafenib area under the curve (AUC) and hepatic function tests (albumin, bilirubin and AST) or creatinine clearance. A significant ($P < 0.001$) reduction in the circulating unbound (to the proteins) fraction of sorafenib has been observed when normal patients have been compared with patients with both liver and renal dysfunction [24].

The most used system to score the liver function is the CP score. Patients with chronic liver disease are classified CP-A to -C, with CP-A accounting for the less compromised function, based on the points scored in five clinical variables: albumin, bilirubin, prothrombin time, ascites and encephalopathy. In general, CP-C patients are poor candidates for clinical research, due to their poor prognosis without liver transplantation.

On the other hand, although CP-A patients represent only a minority of the patients with advanced HCC candidate to systemic treatment, the trials aimed to register new treatments in HCC were restricted to this population [2, 3].

A great debate exists in judging the inclusion in clinical trials of CP-B patients, which represent a relevant proportion of those candidates to a systemic treatment [6, 25].

It has been suggested that in these subjects, death-related cirrhosis prevents from observing the actual antitumour effect of the tested treatments. Therefore, the clinical development of new anticancer therapeutics in HCC should be aimed, though at least initially, to demonstrate the efficacy in the best-available scenario, thus enrolling only CP-A patients. Only when a substantial benefit has been demonstrated in these patients, further investigations in CP-B are warranted [6].

However, one could agree with this process if the initial focus, in the best-available scenario, would not automatically terminate with the approval of the new drug to the general practice. As an example of the latter concept, sorafenib was approved by both Food and Drug Administration (FDA) and European Medicine Agency for the first-line treatment of HCC

patients independently of CP status, based on the results of two phase III trials restricted to CP-A [26]. Interestingly, it has been also said that the approval of sorafenib for HCC patients irrespective of the CP score has left the oncology community in a conundrum: a physician has to face the difficult choice between giving the drug to a CP-B patient without any definitive information regarding the potential benefits and harms to the patient or not to give a drug to the patient, formally a candidate to receive it and lacking any other standard treatment option [27]. Moreover, based on the available data, CP-B patients may gain less survival benefit [28, 29] with higher variability in plasma concentrations of the drug [28] and an increased probability of drug-related serious adverse events [29]. Interestingly, the problem of generalizability has been recognised even at some regulatory levels as the evidence produced by Abou-Alfa et al., albeit derived from an open-label phase II trial, has been used by NICE as a theoretical basis to reject the approval of sorafenib for HCC patients [30]. It has been suggested to perform *ad hoc* trials since evidence from small non-comparative trials conducted in CP-B patients can underestimate as much as phase III trials conducted only in CP-A can overestimate the real benefit gained by 'real world' HCC patients treated with sorafenib.

ethnicity

Ethnicity is a widely accepted risk factor for the development of HCC [31]. In particular, the prevalence of HCC is much higher in the Asian population, whose members are more commonly infected by HBV than in Western ones [32]. However, the existence of differences in mortality and other HCC-related factors between ethnic groups is still controversial [31].

In a retrospective study, Wong et al. have assessed the potential ethnicity based differences in HCC presentation in a cohort of 276 patients, 162 of whom were Asian-Americans and 114 were non-Asian-Americans [31]. Overall, when compared with non-Asian-Americans, Asian-Americans presented a significantly higher incidence of HBV infection history and family history of HCC. Moreover, Asian-Americans had lower CP scores (class A: 62.0% versus 31.4%) and presented with a lower stage of HCC (Okuda staging; I: 43.8% versus 22.8%).

In a wider case–control study ($n = 1040$), conducted at the Stanford University Medical Center, Asian ethnicity was identified in univariate analysis as an independent risk factor for the development of HCC in patients with underlying liver disease [OR, 1.6 versus non-Asian ethnicity; CI 95% (1.2–2.2)], and this result was confirmed in a multivariate model inclusive of several risk factors (age, gender, cirrhosis status, Asian versus non-Asian, AFP ≥50 ng/ml, cumulative cigarette use, heavy alcohol consumption, aetiology of liver diseases and diabetes mellitus) [33]. The existence of differences in the risk and presentation of HCC between Asian and non-Asian populations is further supported by a recent study of tissue array analysis: in samples from Asian patients, more frequent positive staining for p53 (24%) was reported when compared with the American group (9%) [30]. Conversely, a lower frequency of positive staining for MDM2 was observed (2%

versus 26%). These results could lend some support to differences in the molecular pathogenesis of HCC in different populations.

Collectively, the above-mentioned findings suggest that, in order to enhance the robustness of clinical trials, the included population should be stratified according to ethnicity as necessary.

other factors

Other prognostic factors affecting long-term survival and treatment outcomes have been tested, and may therefore potentially represent useful stratification criteria in clinical trials. We briefly report the results of some recent studies addressing this issue.

Cheng et al., in a study with 879 Taiwanian patients from 1993 to 2005, identified a lack of tumour encapsulation, AST values >68 U/l and blood loss >500 ml as independent prognostic factors for disease-free survival (DFS), whereas a lack of tumour encapsulation, AST values >68 U/l, blood loss >500 ml, and serum α-fetoprotein values >200 ng/mL were independent factors impairing overall survival (OS) [34].

A prospective study with 786 HCC patients who received locoregional therapy has investigated the prognostic accuracy of total tumour volume (TTV) [35]. After adjustment for confounding factors, statistical analysis disclosed that patients with TTV 50–200 cm^3 [relative risk (RR): 1.74, $P = 0.009$], 200–500 cm^3 [RR: 2.15, $P = 0.006$] and >500 cm^3 [RR: 3.92, $P < 0.001$] presented a higher mortality risk when compared with patients with TTV <10 cm^3, suggesting that the TTV could represent a feasible prognostic predictor and a potentially useful parameter for mortality risk stratification.

Lastly, Kaseb et al. have hypothesised that insulin-like growth factor-1 (IGF-1), a growth factor mainly produced by the liver, could correlate with patients' survival and hence improve the prognostic ability of the CLIP score [36]. From the analysis of IGF-1 concentrations and clinic-pathologic parameters of 288 HCC patients, the authors observed that IGF-1 was significantly correlated with the clinic-pathologic features: the CLIP criteria can, therefore, be extended to include IGF-1 (I-CLIP) to improve prognostic stratification of HCC patients in clinical trials.

end points

In oncology practice, the primary goals of the clinicians are to prolong the survival and alleviate the suffering of their patients. As a consequence, demonstrating a meaningful benefit in both survival and symptom relief is the main answer clinicians expect to find, at least, in late-stage clinical research and should be strongly considered in choosing clinical trials objectives.

In the case of trials with HCC patients, several considerations need to be made. The dependency of HCC patients' survival upon liver disease rather than tumour progression or treatment failure, the crucial role of liver function in drug metabolism and tolerability and, finally, the lack of therapeutic alternatives after the first line form the background that has to be considered in planning clinical research in HCC.

Arguably, liver function is the most important factor conditioning the choice of the end points for early-phase trials. As cited above, a recent phase I trial recommended different doses of sorafenib according to different degrees of liver function [24]. The definition of the recommended dose based only on the maximum-tolerated dose could be misleading in patients with impaired liver function. Since both toxicity and activity ultimately depend on circulating drug levels, though at different levels respectively, it could be easily concluded that pharmacokinetic end points should be necessarily evaluated in phase I trials and that the recommended dose for phase II should be based on accurate PK/PD modelling rather than only on toxicity. However, toxicity still has to be taken into account evaluating a new treatment in phase I trials, especially in cases like sorafenib where a PK/organ function relationship seems non-significant [24, 28, 37] in contrast to the toxicity–organ function relationship.

In late-stage clinical trials, OS, that is the time from the start of the treatment (i.e. randomisation in the case of randomised trials) to the death of the patient, is undoubtedly the optimal parameter to estimate the benefit and definitely will provide clinicians with the answer to the question on how the treatment prolongs the survival of treated patients. The importance of OS as treatment goal is evident to patients and physicians. Both, though mostly the patients have declared to be willing to face adverse effects of treatment balanced by an even small survival benefit [38]. From a methodological perspective, OS is as close to the ideal end point as it can be considered easy to be measured in an error-free and unbiased way [39, 40], it is objective and represents an unquestionable benefit for the patients. Therefore, it is considered the most important end point. In fact, FDA has considered OS as a direct, universally accepted measure of treatment benefit and preferred clinical trial end point when possible [41]. However, the OS as end point of clinical trials can be hindered by some disadvantages. The availability of effective therapies administered after the trial impairs the ability of OS to catch a real benefit gain in a clinical trial [42], although this is not the case in advanced HCC where no effective second-line treatments exist in clinical practice. Similarly, the crossover within a comparative trial can obscure a potential benefit. Other 'OS as end point' disadvantages comprise the large number of patients and the long follow-up required that recently hampered its more frequent use, at least for patients without advanced disease. Finally and greatly interesting for HCC trials, it has been argued that OS is not an optimal end point in HCC because it is conditioned by competing causes of death [6]. Theoretically, the latter point is not an obstacle to the use of OS as principal end point and should be only taken into account during the analysis of survival data in clinical trials. Even in a tumour type where the prognosis is mostly governed by other causes, like the liver disease in the HCC, prolonging survival of patients remains the main goal of the oncology treatment. One could argue that it is difficult for OS to catch a potential benefit offered by a treatment, due to an even modest rate of intercurrent deaths [43]. At the same time one could conclude that whether the benefit derived from the anticancer treatment is not of such relevance to patient survival (since it is not even reasonably possible to catch it) as

the tremendous impact of intercurrent deaths, treating the tumour is not as worthwhile as treating the competing causes of death (i.e. liver disease in HCC).

However, an analysis of competing risks of death can be successfully used to better estimate the benefit of a treatment in the case of alternative but plausible causes of death (i.e. liver failure). Notably, this type of analysis will require an unrealistic larger sample size than the OS analysis to permit an acceptable power.

Quality-of-life analysis (QoL) and Time to Symptomatic Progression require particular mention among the potential end points. In particular, they respond to the other fundamental question the clinician faces considering a treatment implementation: 'how will the treatment alleviate the suffering of my patient?'.

However, differently from the OS, QoL and in general the impact of a treatment on the symptoms burden is not easily measurable. First, the questionnaires generally used to evaluate QoL are not validated for patients with HCC. FACT-Hep [44] and EORTC QLQ-C30 [45] are the most used and, at least the first, has been developed for patients with hepatobiliary malignancies. Second it is very difficult to judge when worsening of symptoms is due to tumour progression rather than to the underlying cirrhosis [6].

The drawbacks, described above, of OS as the primary end point have prompted the recent, more frequent use of efficacy end points based on tumour assessments [6, 42].

Progression-free survival (PFS) and time to progression (TTP) are the most used in advanced setting, while DFS and time to recurrence (TTR) are used in adjuvant setting. The main difference among the 'survival' end point and 'time to' end points is the type of the events are required for the analysis. In case of DFS and PFS the events required include both progression/recurrence and death from any cause. In TTP and TTR the only event of interest is the progression or recurrence whilst the patients dead at a certain time are censored at that time. PFS and TTP are often used invariantly in recent literature. However, PFS should be preferred from a regulatory point of view due to its ability to catch fatal events due to untoward treatment effects [46]. Overall PFS and TTP might be attractive compared with OS because earlier to be registered, requiring a lower number of patients and not being prone to confounding factors as subsequent treatments administered after the trial. Nevertheless, they can just indirectly estimate a real clinical benefit derived from a treatment [47]. In contrast, on a theoretical basis, a trial powered to catch a meaningful OS benefit is also well powered to catch a PFS or TTP benefit whilst the opposite is never true. Moreover, at least PFS, in hepatic-dysfunctioned HCC patients where the life expectancy is often no longer than the time on treatment due to the liver disease, is as prone as OS to the confounding competing causes of death. Interestingly, the results of a recent survey conducted by French oncologists and methodologists have shown how the 'optimal' alternative end points can vary according to the stage of HCC [48]. This survey was extended to several types of gastrointestinal cancers: with respect to HCC, DFS, PFS, local control and RR were identified as preferred alternative end points in patients with

early disease. In case of advanced HCC and metastatic disease, the preferred end point was PFS; the potential role of Quality of Life (QoL) was also suggested, in line with some previous observations and expert opinions [49, 50].

Moreover, all the tumour assessment based end points, comprising Response Rate (RR) suffer from other issues. In fact, they are largely dependent on tumour assessment interval and on the tumour assessment criteria.

timing of radiological assessment

To minimize the risk of the first potential bias, the re-evaluation between the two arms has to be synchronized and an interval of 6–8 weeks, symmetrically in both treatment arms [6] could be recommended.

criteria for the assessment of radiological response

Historically, tumour response to treatment was measured, from a radiological point of view, by the World Health Organization (WHO) criteria [51], and afterwards according to the Response Evaluation Criteria in Solid Tumours (RECIST) [52]. The WHO and RECIST criteria define standard methods for the conversion of radiology images into measurable values to assess the response to therapy in terms of tumour size [53]. Criteria for evaluating lesions have varied according to different versions of the RECIST criteria [52, 54].

However, both WHO and RECIST criteria were developed in the era of cytotoxics, thus considering only tumour shrinkage due to cell death as a measure of antitumour activity [53]. Such a measure can be misleading when applied to new targeted anti-cancer drugs with a different mechanism of action, such as the inhibition of angiogenesis or to other therapeutic interventions, such as transarterial chemoembolization (TACE) [52, 53]. These therapies may indeed determine tumour devascularisation and necrosis, not being necessarily associated with a reduction in size. In the landmark phase III SHARP (Sorafenib Hepatocellular Carcinoma Assessment Randomised Protocol) study, which assessed the efficacy and safety of sorafenib, a poor correlation between the clinical benefit in terms of survival and the RR according to the RECIST criteria was shown [2]. On these bases, the AASLD proposed a formal amendment of the RECIST criteria, which takes into consideration the degree of tumour arterial enhancement: the modified RECIST criteria (mRECIST) [53]. In addition, the European Association for the Study of Liver Disease (EASL) has defined different criteria—the EASL criteria—for the assessment of response [55]. Table 1 summarizes the definitions of response according to the EASL, RECIST 1.1 and mRECIST criteria [56].

The prognostic abilities of these criteria for the assessment of response have been compared in different studies [56, 57]. Gilmore et al. determined the tumour response—using RECIST 1.1, EASL and mRECIST criteria—in 83 HCC patients treated with TACE as a palliative therapy [56]. Overall, a good correlation between EASL and mRECIST was shown, with overall response rates of 58% and 57%, and target lesion responses of 74% and 73%, respectively. Conversely, a poor

Table 1. Different radiological criteria (EASL, RECIST 1.1, and mRECIST) for the evaluation of response to anticancer treatments (reproduced from [56], with permission)

	EASL	RECIST 1.1	mRECIST
Complete response (CR)	Disappearance of any intratumoural arterial enhancement in all measurable arterially enhancing liver lesions.	Disappearance of all target lesions (up to two measurable liver lesions).	Disappearance of any intratumoural arterial enhancement in all target lesions (up to two measurable liver lesions).
Partial response (PR)	At least a 50% decrease in the sum of the product of bi-dimensional diameters of viable (enhancement in the arterial phase) target lesions, taking as reference the baseline sum of the diameters of target lesions.	At least a 30% decrease in the sum of the greatest unidimensional diameters of target lesions, taking as reference the baseline sum of the diameters of target lesions.	At least a 30% decrease in the sum of unidimensional diameters of viable (enhancement in the arterial phase) target lesions, taking as reference the baseline sum of the diameters of target lesions.
Stable disease (SD)	Any cases that do not qualify for either partial response or progressive disease.	Any cases that do not qualify for either partial response or progressive disease.	Any cases that do not qualify for either partial response or progressive disease.
Progressive disease (PD)	An increase of at least 25% in the sum of the diameters of viable (enhancing) target lesions, taking as reference the smallest sum of the diameters of viable (enhancing) target lesions recorded since treatment started.	An increase of at least 20% in the sum of the diameters of target lesions, taking as reference the smallest sum of the diameters of target lesions recorded since treatment started.	An increase of at least 20% in the sum of the diameters of viable (enhancing) target lesions, taking as reference the smallest sum of the diameters of viable (enhancing) target lesions recorded since treatment started.

correlation for RECIST 1.1 criteria was reported, with overall and target response rates of 7%. In addition, statistical analysis disclosed a significant association between OS and overall response according to both the EASL (44% risk reduction of death) and mRECIST (42%) criteria, while there was no significant association between OS and RECIST 1.1 response. These results suggested that EASL and mRECIST overall response rates may be predictors of survival and these criteria might be used in preference to RECIST 1.1 criteria [56].

In another study, Edeline et al. retrospectively analysed 53 patients with advanced HCC treated with sorafenib, and compared the response rates observed with the mRECIST and the RECIST 1.1 criteria [57]. Patients who achieved an overall response according to mRECIST criteria had a longer OS than non-responding patients (median OS, 18 months and 8 months, respectively). In addition, among the 42 patients who achieved SD according to RECIST, OS differed depending upon mRECIST tumour response, with a median OS of 17 months, 10 months, and 4 months for patients who achieved an overall response, a stable disease, or a progressive disease. The authors concluded that mRECIST might be used for the standard assessment of treatment activity in HCC patients in clinical trials and clinical practice [57]. In addition, and differing from what reported for locoregional therapies due to the peculiar mechanism of action of targeted therapies, the differences observed in vascularisation could reflect a real biological change in the tumour, associated with clinically meaningful implications for the patient even if the threshold for response is not met [57].

Finally, the relative value of mRECIST when compared with mWHO criteria has been evaluated by Finn et al. at the ASCO 2010 annual meeting [58]. Using the data from a phase II trial with brivanib, a multikinase inhibitor, they found a relevant discrepancy in the responses between mWHO (5 of 101) and mRECIST (19 of 101) criteria. Moreover, 31 of 51 patients

experiencing a PD according to mWHO had SD according to mRECIST and may have interrupted the treatment too early.

the importance of a radiological assessment

In recent years, the importance of a proper radiological assessment in both the diagnosis and evaluation of treatments for HCC, both in experimental trials and in clinical practice, has been stressed in several publications [59–61]. Presently, the imaging techniques commonly used for the diagnosis of HCC are ultrasound (US), computed tomography (CT) and MRI [59].

It is widely accepted that HCC emerges as a small nodule composed of well-differentiated hepatocytes, and then progresses at a heterogeneous rate into a larger nodule [61]. At US, most small nodules appear hypoechoic, but some are hyperechogenic due to a microsteatosis that may disappear when tumour progresses. Major angiogenesis occurs between 10 and 20 mm, and most HCC >20 mm are intensely hypervascular. Therefore, an intense contrast uptake in the arterial phase followed by contrast washout in the delayed venous phase should be used at dynamic imaging by CT/MRI [61]. Arterial uptake may also be recognized by contrast enhanced US (CEUS); however, CT or MRI currently represents the most reliable method for the diagnosis of HCC [61].

Imaging techniques do play a central role also in the evaluation of treatment efficacy. Follow-up examination with CT or MRI is recommended every 3–6 months after surgical resection, in order to identify possible recurrences [60, 61]. In patients treated with locoregional therapies, the efficacy of imaging techniques to assess initial treatment success differs with the tumour size [61]. In HCC <2 cm, any early assessment after therapy may be misleading due to inflammatory changes. In larger tumours, periprocedural

CEUS could be used to identify the non-ablated areas, while CT or MRI are effective for follow-up monitoring, usually at 1 or 2 months after the procedure, every 3 months during the first 2 years, and every 4–6 months thereafter [61]. With respect to sorafenib treatment, a recent study has evaluated the role of MRI in the assessment of tumour response [62]. In detail, MRI signal patterns were assessed in 21 advanced HCC patients, at baseline and at short-term intervals thereafter (survey time 2–65 weeks). Signal abnormalities were disclosed by T1WI and T2WI in 15 of 21 patients. The predominant tumour signal change was hyperintensity on both T1WI and T2WI. Of note, most patients developed MRI signal changes within 4 weeks of therapy, while two non-responders did not show any signal alteration at follow-up. These findings, albeit preliminary, could suggest that early MRI-based evaluation could play a role in the evaluation of the efficacy of sorafenib treatment.

On these bases, US, CT and MRI should be considered valuable and non-invasive imaging modalities for diagnosis of HCC and the evaluation of treatment response in both a clinical and an experimental setting. However, at present no imaging technique is able to effectively diagnose small (<1 cm) arterial enhancing lesions [59]; it has been suggested that further improvement of imaging technologies—including functional imaging such as elastography, perfusion imaging and diffusion imaging—together with the development of new contrast media has the potential to improve the detection and characterization of these small tumours [59].

study design

As noted above, most of the recommendations in HCC made even in the most updated guidelines are based on low strength evidence. Similarly, most of the conclusions regarding the efficacy and toxicity of particular treatments in particular populations [29, 63] are drawn from weak evidence.

Large, well-conducted, observational, studies might provide complementary results to those collected in randomised clinical trials (RCTs), as they assess the effectiveness of a given treatment in patients encountered in day-to-day clinical practice, potentially affected by different co-morbidities or taking multiple treatments. In addition, well-designed observational studies can help identify clinically relevant differences among therapeutic options and collect information on long-term drug effectiveness and safety [64]. Even when well conducted they can nevertheless only suggest hypotheses, they do not provide strong evidence for treatment benefit.

Well-designed and conducted phase III RCTs represent the strongest source of evidence in clinical research. Several factors play a role in making the design of such trials difficult in HCC patients. In particular, as detailed above heterogeneity of HCC patients due to profound differences in prognosis, related or not to the tumour (disease extent and liver function), is one of them. Although homogeneity in disease extent should be warranted in the trial in order to allow proper comparison, stratification before randomisation, i.e. according to the CP score and BCLC stage, can resolve most of this heterogeneity without compromising the study power [25]. In this context, randomisation is the only way to ensure the unbiased

distribution of all the known and unknown potential confounding factors between the two arms. In fact, the use of external controls (indirect comparison, historical controls or large series) is restricted by the inability to control bias [65] even when the comparison is made between apparently very close populations [66]. As discussed above, this limits the role of this kind of evidence, although it can provide useful information to draw any informed conclusion or recommendation to clinical practice.

concluding remarks

In 2012, HCC is still a major health problem. Although HCC is only the sixth most common tumour, its mortality is disproportionately high (third after lung and stomach) when compared with similarly prevalent tumours. Prevention tools and curative treatments are available only for a minority of patients with HBV or non-infection-related HCC and with local disease, respectively. Despite efforts have been made to foster clinical research in HCC, dismal results have been obtained until now and only one drug is available to patients with advanced disease. The major problem in pursuing the clinical benefit demonstration for HCC patients is that most of them suffer from underlying liver impairment potentially affecting both tolerability and efficacy of the treatment. An *ad hoc* pathway of clinical development for new therapies should be implemented with early trials conducted separately in different populations of patients according to their liver function (Figure 1). Pharmacodynamic/pharmacokinetic end points should be used in phase I trials to define the recommended dose for subsequent trials. However, phase I trials still have to consider toxicity as an important end point and should be designed to report it thoroughly. In the late-stage trials, OS should be always considered as the most important end point, and preferred as the primary one. Alternative end points should also be considered as secondary and reported along with OS. The competing causes of death, indeed, do not impair the value of OS as treatment goal, nor warrant an adequate surrogacy power to alternative end points such as PFS and TTP. Moreover, these end points suffer from potential bias related to the response evaluation, and the
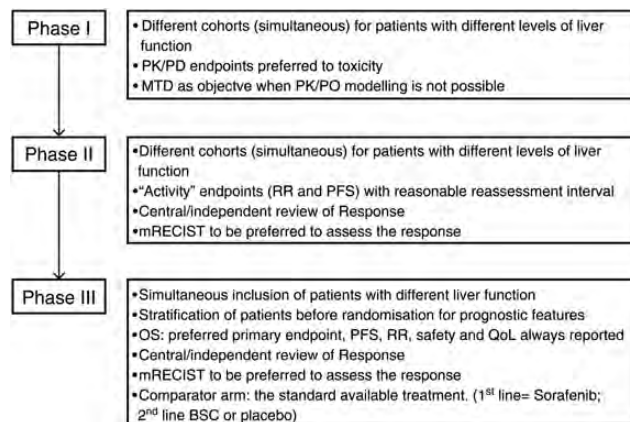

Figure 1. Development pathway for clinical development of new treatments in HCC.

timing of tumour assessment. In fact, when reporting PFS and TTP, the method of assessment and the assessment review (central or independent or local) should also be explicitly reported.

Finally, randomised trials with adequate randomisation techniques, stratification before randomisation and fair comparison are needed to address the clinical problem of physicians and patients facing up with HCC. Observational studies are useful to collect information on a large scale, but in the absence of evidence coming from randomised trials they should be avoided and their resources should be used to build robust evidence.

All the phases of investigations in HCC should proceed contemporaneously. The urgency to gain favourable results in a small proportion of patients is not worthwhile if it ultimately leaves the clinical community in the doubt to treat the entire population of patients without enough information.

funding

disclosure

The authors declare no conflicts of interest directly relevant to this study.

references

1. Jemal A, Bray F, Center MM et al. Global cancer statistics. CA Cancer J Clin 2011; 61(2): 69–90.

2. Llovet JM, Ricci S, Mazzaferro V et al. SHARP Investigators Study Group. Sorafenib in advanced hepatocellular carcinoma. N Engl J Med 2008; 359(4): 378–390.

3. Cheng AL, Kang YK, Chen Z 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(1): 25–34.

4. Di Maio M, De Maio E, Perrone F et al. Hepatocellular carcinoma: systemic treatments. J Clin Gastroenterol. 2002; 35(5 Suppl 2): S109–S114.

5. European Association For The Study Of The Liver; European Organisation For Research And Treatment Of Cancer. EASL-EORTC clinical practice guidelines: management of hepatocellular carcinoma. J Hepatol 2012; 56(4): 908–943.

6. Llovet JM, Di Bisceglie AM, Bruix J et al. Panel of Experts in HCC-Design Clinical Trials. Design and endpoints of clinical trials in hepatocellular carcinoma. J Natl Cancer Inst 2008; 100(10): 698–711.

7. de Bono JS, Ashworth A. Translating cancer research into targeted therapeutics. Nature 2010; 467(7315): 543–549.

8. Rosell R, Carcereny E, Gervais R et al. Spanish Lung Cancer Group in collaboration with Groupe Français de Pneumo-Cancérologie and Associazione Italiana Oncologia Toracica. Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. Lancet Oncol 2012; 13(3): 239–246.

9. Chapman PB, Hauschild A, Robert C et al. BRIM-3 Study Group. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. N Engl J Med 2011; 364(26): 2507–2516.

10. Gelmon KA, Tischkowitz M, Mackay H et al. Olaparib in patients with recurrent high-grade serous or poorly differentiated ovarian carcinoma or triple-negative breast cancer: a phase 2, multicentre, open-label, non-randomised study. Lancet Oncol 2011; 12(9): 852–861.

11. Rodón J, Saura C, Dienstmann R et al. Molecular prescreening to select patient population in early clinical trials. Nat Rev Clin Oncol 2012; 9(6): 359–366.

12. Von Hoff DD, Stephenson JJ, Jr, Rosen P et al. Pilot study using molecular profiling of patients' tumors to find potential targets and select treatments for their refractory cancers. J Clin Oncol 2010; 28(33): 4877–4883.

13. Bruix J, Sherman M; Practice Guidelines Committee, American Association for the Study of Liver Diseases. Management of hepatocellular carcinoma. Hepatology 2005; 42(5): 1208–1236.

14. Stigliano R, Marelli L, Yu D et al. Seeding following percutaneous diagnostic and therapeutic approaches for hepatocellular carcinoma. What is the risk and the outcome? Seeding risk for percutaneous approach of HCC. Cancer Treat Rev 2007; 33(5): 437–447.

15. Buyse M, Sargent DJ, Grothey A et al. Biomarkers and surrogate end points—the challenge of statistical validation. Nat Rev Clin Oncol 2010; 7(6): 309–317.

16. Daniele B, Perrone F. Staging for liver cancer. Clin Liver Dis 2005; 9(2): 213–223, vi.

17. Llovet JM, Bruix C, Bruix J. Prognosis of hepatocellular carcinoma: the BCLC staging classification. Semin Liver Dis 1999; 19(3): 329–338.

18. A new prognostic system for hepatocellular carcinoma: a retrospective study of 435 patients: the Cancer of the Liver Italian Program (CLIP) investigators. Hepatology 1998; 28(3): 751–755.

19. Prospective validation of the CLIP score: a new prognostic system for patients with cirrhosis and hepatocellular carcinoma. The Cancer of the Liver Italian Program (CLIP) Investigators. Hepatology 2000; 31(4): 840–845.

20. Leung TW, Tang AM, Zee B et al. Construction of the Chinese University Prognostic Index for hepatocellular carcinoma and comparison with the TNM staging system, the Okuda staging system, and the Cancer of the Liver Italian Program staging system: a study based on 926 patients. Cancer 2002; 94(6): 1760–1769.

21. Durand F, Valla D. Assessment of the prognosis of cirrhosis: Child-Pugh versus MELD. J Hepatol 2005; 42(Suppl): S100–107.

22. Shetty K, Rybicki L, Carey WD. The Child-Pugh classification as a prognostic indicator for survival in primary sclerosing cholangitis. Hepatology 1997; 25(5): 1049–1053.

23. Peck-Radosavljevic M, Greten TF, Lammer J et al. Consensus on the current use of sorafenib for the treatment of hepatocellular carcinoma. Eur J Gastroenterol Hepatol 2010; 22(4): 391–398.

24. Miller AA, Murry DJ, Owzar K et al. Phase I and pharmacokinetic study of sorafenib in patients with hepatic or renal dysfunction: CALGB 60301. J Clin Oncol 2009; 27(11): 1800–1805.

25. Di Maio M, Daniele B, Gallo C et al. Re: Design and endpoints of clinical trials in hepatocellular carcinoma. J Natl Cancer Inst. 2008; 100(21): 1557. author reply 1557–8.

26. Di Maio M, Daniele B, Perrone F. Targeted therapies: Role of sorafenib in HCC patients with compromised liver function. Nat Rev Clin Oncol. 2009; 6(9): 505–506.

27. Zhu AX, Clark JW. Commentary: Sorafenib use in patients with advanced hepatocellular carcinoma and underlying Child-Pugh B cirrhosis: evidence and controversy. Oncologist. 2009; 14(1): 67–69.

28. Abou-Alfa GK, Amadori D, Santoro A et al. Safety and efficacy of sorafenib in patients with hepatocellular carcinoma (HCC) and Child-Pugh A versus B cirrhosis. Gastrointest Cancer Res 2011; 4(2): 40–44.

29. Marrero JA, Lencioni R, Kudo M et al. Global investigation of therapeutic decisions in hepatocellular carcinoma and of its treatment with sorafenib (GIDEON) second interim analysis in more than 1,500 patients: clinical findings in patients with liver dysfunction. J Clin Oncol 2011; 29(suppl); abstr 4001.

30. Hepatocellular carcinoma (advanced and metastatic)—sorafenib (first line) (TA189). http://guidance.nice.org.uk/TA189.

31. Wong PY, Xia V, Imagawa DK et al. Clinical presentation of hepatocellular carcinoma (HCC) in Asian-Americans versus non-Asian-Americans. J Immigr Minor Health 2011; 13(5): 842–848.

32. Bosch FX, Ribes J, Cleries R et al. Epidemiology of hepatocellular carcinoma. Clin Liver Dis 2005; 9: 191–211.

33. Ha NB, Ha NB, Ahmed A et al. Risk factors for hepatocellular carcinoma in patients with chronic liver disease: a case-control study. Cancer Causes Control. 2012; 23: 455–462.

34. Cheng CH, Lee CF, Wu TH et al. Evaluation of the new AJCC staging system for resectable hepatocellular carcinoma. World J Surg Oncol 2011; 9: 114.

35. Huo TI, Hsu CY, Huang YH et al. Prognostic prediction across a gradient of total tumor volume in patients with hepatocellular carcinoma undergoing locoregional therapy. BMC Gastroenterol 2010; 10: 146.

36. Kaseb AO, Abbruzzese JL, Vauthey JN et al. I-CLIP: improved stratification of advanced hepatocellular carcinoma patients by integrating plasma IGF-1 into CLIP score. Oncology 2011; 80(5–6): 373–381.

37. Jain L, Woo S, Gardner ER et al. Population pharmacokinetic analysis of sorafenib in patients with solid tumours. Br J Clin Pharmacol 2011; 72(2): 294–305.

38. Bremnes RM, Andersen K, Wist EA. Cancer patients, doctors and nurses vary in their willingness to undertake cancer chemotherapy. Eur J Cancer 1995; 31A (12): 1955–1959.

39. Pazdur R. Endpoints for assessing drug activity in clinical trials. Oncologist 2008; 13(Suppl 2): 19–21.

40. Sargent D. General and statistical hierarchy of appropriate biologic endpoints. Oncology (Williston Park) 2006; 20(6 Suppl 5): 5–9.

41. Johnson JR, Temple R. Food and Drug Administration requirements for approval of new anticancer drugs. Cancer Treat Rep. 1985; 69(10): 1155–1159.

42. Buyse M, Sargent DJ, Saad ED. Survival is not a good outcome for randomized trials with effective subsequent therapies. J Clin Oncol 2011; 29(35): 4719–4720. author reply 4720–1.

43. Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. Stat Med 2007; 26(11): 2389–2430.

44. Heffernan N, Cella D, Webster K 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(9): 2229–2239.

45. Blazeby JM, Currie E, Zee BC et al. EORTC Quality of Life Group. Development of a questionnaire module to supplement the EORTC QLQ-C30 to assess quality of life in patients with hepatocellular carcinoma, the EORTC QLQ-HCC18. Eur J Cancer 2004; 40(16): 2439–2444.

46. Johnson JR, Williams G, Pazdur R. End points and United States Food and Drug Administration approval of oncology drugs. J Clin Oncol 2003; 21(7): 1404–1411.

47. Fleming TR, Rothmann MD, Lu HL. Issues in using progression-free survival when evaluating oncology products. J Clin Oncol 2009; 27(17): 2874–2880.

48. Methy N, Bedenne L, Bonnetain F. Surrogate endpoints for overall survival in digestive oncology trials: which candidates? A questionnaires survey among clinicians and methodologists. BMC Cancer 2010; 10: 277.

49. Spinzi G, Paggi S. Sorafenib in advanced hepatocellular carcinoma. N Engl J Med 2008; 359(23): 2497–2498. author reply 2498–9.

50. Spinzi G, Terreni N. Health-related quality of life and sorafenib. Hepatology 2010; 52(4): 1523.

51. Miller AB, Hoogstraten B, Staquet M et al. Reporting results of cancer treatment. Cancer 1981; 47(1): 207–214.

52. Therasse P, Arbuck SG, Eisenhauer EA et al. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. J Natl Cancer Inst 2000; 92(3): 205–216.

53. Lencioni R, Llovet JM. Modified RECIST (mRECIST) assessment for hepatocellular carcinoma. Semin Liver Dis 2010; 30(1): 52–60.

54. Eisenhauer EA, Therasse P, Bogaerts J et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer 2009; 45(2): 228–247.

55. Bruix J, Sherman M, Llovet JM et al. Clinical management of hepatocellular carcinoma. Conclusions of the Barcelona-2000 EASL conference. European Association for the Study of the Liver. J Hepatol 2001; 35: 421–430.

56. Gillmore R, Stuart S, Kirkwood A et al. EASL and mRECIST responses are independent prognostic factors for survival in hepatocellular cancer patients treated with transarterial embolization. J Hepatol 2011; 55(6): 1309–1316.

57. Edeline J, Boucher E, Rolland Y 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 (1): 147–156.

58. Finn RS, Raoul J, Manekas D et al. Optimal assessment of treatment benefit with targeted agents for hepatocellular carcinoma (HCC): an analysis of brivanib phase II data. J Clin Oncol 2010; 28(suppl); abstr 4096, 15s.

59. Choi BI, Lee JM. Advancement in HCC imaging: diagnosis, staging and treatment efficacy assessments: imaging diagnosis and staging of hepatocellular carcinoma. J Hepatobiliary Pancreat Sci 2010; 17(4): 369–373.

60. Vilgrain V. Advancement in HCC imaging: diagnosis, staging and treatment efficacy assessments: hepatocellular carcinoma: imaging in assessing treatment efficacy. J Hepatobiliary Pancreat Sci 2010; 17(4): 374–379.

61. Bruix J, Reig M, Rimola J et al. Clinical decision making and research in hepatocellular carcinoma: pivotal role of imaging techniques. Hepatology 2011; 54(6): 2238–2244.

62. Horger M, Lauer UM, Schraml C et al. Early MRI response monitoring of patients with advanced hepatocellular carcinoma under treatment with the multikinase inhibitor sorafenib. BMC Cancer 2009; 9(1): 208.

63. Iavarone M, Cabibbo G, Piscaglia F et al. SOFIA (SOraFenib Italian Assessment) study group. Field-practice study of sorafenib therapy for hepatocellular carcinoma: a prospective multicenter study in Italy. Hepatology 2011; 54(6): 2055–2063.

64. Silverman SL. From randomized controlled trials to observational studies. Am J Med. 2009; 122(2): 114–120.

65. COMMITTEE FOR MEDICINAL PRODUCTS FOR HUMAN USE (CHMP). Guideline on the evaluation of anticancer medicinal products in man. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/12/WC500017748.pdf.

66. Sangro B, Carpanese L, Cianni R et al. Assessment of radioembolization among patients who met the inclusion criteria for Sorafenib Hepatocellular Carcinoma Assessment Randomized Protocol (SHARP). J Clin Oncol 2012; 30 (suppl; abstr e14654).