COMMENTARY

Design and Endpoints of Clinical Trials in Hepatocellular Carcinoma

Josep M. Llovet, Adrian M. Di Bisceglie, Jordi Bruix, Barnett S. Kramer, Riccardo Lencioni, Andrew X. Zhu, Morris Sherman, Myron Schwartz, Michael Lotze, Jayant Talwalkar, Gregory J. Gores; for the Panel of Experts in HCC-Design Clinical Trials

The design of clinical trials in hepatocellular carcinoma (HCC) is complex because many patients have concurrent liver disease, which can confound the assessment of clinical benefit. There is an urgent need for high-quality trials in this disease. An expert panel was convened by the American Association for the Study of Liver Diseases to develop guidelines that provide a common framework for designing trials to facilitate comparability of results. According to these guidelines, randomized phase 2 trials with a time-to-event primary endpoint, such as time to progression, are pivotal in clinical research on HCC. Survival remains the main endpoint to measure effectiveness in phase 3 studies, whereas time to recurrence is proposed as an appropriate endpoint in the adjuvant setting. Because progression-free survival and disease-free survival are composite endpoints, they are more vulnerable than others in HCC clinical studies and may not be able to capture clinical benefits. Selection of the target population should be based on the Barcelona Clinic Liver Cancer staging system. New drugs should be tested in patients with well-preserved liver function (Child–Pugh A class). Patients assigned to the control arm should receive standard-of-care therapy, that is, chemoembolization for patients with intermediate-stage disease and sorafenib for patients with advanced-stage disease. Further research is needed to incorporate biomarkers and molecular imaging into clinical research in HCC. These surrogate markers may help to enrich study populations and maximize the cost–benefit ratio of trial execution. Design and conduct of phase 3 trials should be coordinated by centers with appropriate expertise in HCC.

Hepatocellular carcinoma (HCC) is a major health problem worldwide. It is the sixth most common neoplasm in the world, with more than half a million new cases annually (1), and the main cause of death among cirrhotic patients (2). The incidence of HCC is increasing in the United States and Europe and within two decades is expected to equal that currently experienced in Japan (3). Chronic infection with hepatitis B virus is the predominant risk factor for HCC in Southeast Asia and Africa, and chronic infection with hepatitis C virus is the predominant risk factor for HCC in Western countries and Japan. Potentially curative treatments, mostly surgical in nature (ie, resection and liver transplantation), are applied to 30%–40% of patients in the West but a smaller proportion of patients in Asia (4). Few medical interventions have been thoroughly tested in HCC, in contrast with the many tested in other highly prevalent cancers, such as lung, breast, and colorectal cancer. In fact, thousands of randomized controlled trials (RCTs) evaluating therapeutic interventions have been published in these cancers, compared with only approximately 80 in HCC (5,6). Two-thirds of these studies had less than optimal study designs and methodology (5,6).

Several factors contribute to the difficulties in designing clinical trials in HCC. Inclusion criteria are frequently heterogeneous because of the lack of a consensus regarding a universal standardized classification, and this heterogeneity has made it difficult to compare, analyze, and interpret results of reported studies. Standardization of diagnostic criteria, stratification factors, and appropriate primary and surrogate endpoint are crucial for the advancement of clinical research in HCC and for capturing benefits in health outcomes that might be achieved with new interventions.

The advent of molecular targeted therapies in oncology has also challenged the use of conventional endpoints (eg, response rate) in phase 2 and 3 trials in HCC. Given these unmet needs for designing clinical trials in HCC, a group of experts was convened by the American Association for the Study of Liver Diseases (AASLD) in December 2006 to participate in a special "Endpoints" conference. The panel developed a set of guidelines that are aimed at providing a framework for the design of clinical trials by integrating
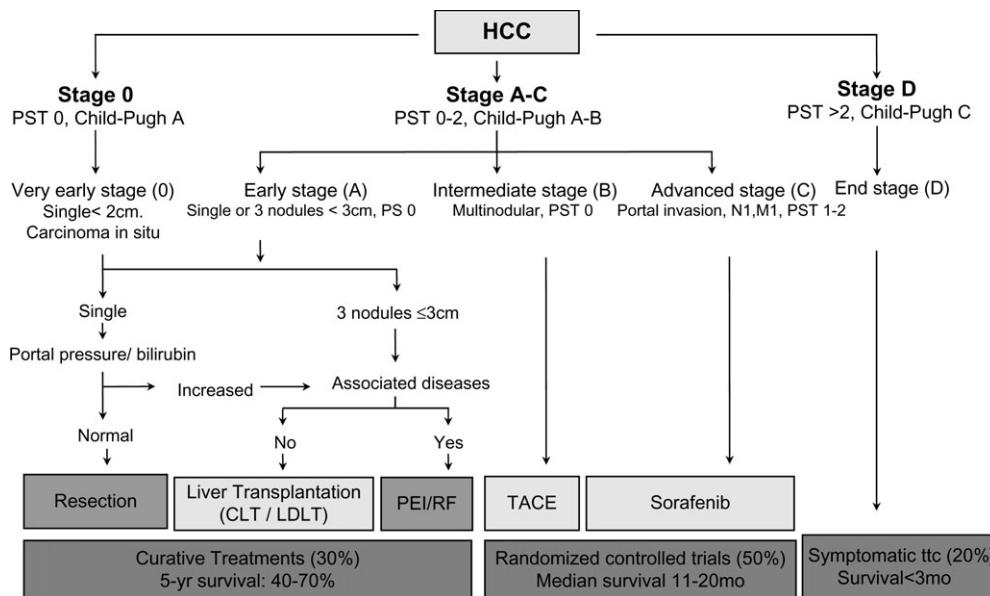
Figure 1. Barcelona Clinic Liver Cancer (BCLC) staging classification and treatment schedule. Patients with very early hepatocellular carcinoma (HCC) (stage 0) are optimal candidates for resection. Patients with early HCC (stage A) are candidates for radical therapy (resection, liver transplantation [LT], or local ablation via percutaneous ethanol injection [PEI] or radiofrequency [RF] ablation). Patients with intermediate HCC (stage B) benefit from transarterial chemoembolization (TACE). Patients with advanced HCC, defined as presence of macroscopic vascular invasion, extrahepatic spread, or cancer-related symptoms (Eastern Cooperative Oncology Group performance status 1 or 2) (stage C), benefit from sorafenib. Patients with end-stage disease (stage D) will receive symptomatic treatment. Treatment strategy will transition from one stage to another on treatment failure or contraindications for the procedures. Modified from Llovet et al. (4).

knowledge about the standard of care in the disease with state-of-the-art knowledge in trial methodology. In this commentary, we summarize the natural history and standard treatments in HCC, review endpoints in targeted therapies in oncology, and provide the panel's recommendations on trial design, endpoints, and selection of study population and control arms in HCC. We also point out the main challenges of conducting trials in this disease.

Natural History, Prognosis, and Standard Treatments for Hepatocellular Carcinoma

An understanding of the natural history and prognostic factors in patients with HCC is essential for the design of clinical trials, analysis of statistical power, and appropriate patient stratification. The natural history and outcomes of HCC according to treatment have been previously established by the AASLD (7) and the European Association for the Study of the Liver (EASL) practice guidelines (8). Those guidelines also established noninvasive diagnostic criteria for HCC that the expert panel accepted for trial design. Once a diagnosis has been made, patient prognosis will vary according to disease stage and treatment received. The main prognostic factors are related to tumor status (defined by the number and size of nodules, the presence or absence of vascular invasion, and the presence or absence of extrahepatic spread), liver function (defined by Child–Pugh class [Supplementary Table 1, available online], serum bilirubin and albumin levels, and portal hypertension), and general health status (defined by Eastern Cooperative Oncology Group [ECOG] classification [Supplementary Table 2, available online] and presence of symptoms). Etiology is not an independent prognostic factor.

Several classification systems are available for HCC. The Barcelona Clinic Liver Cancer (BCLC) classification (Supplementary Table 3, available online) has emerged during recent years as the standard classification that is used for trial design and clinical management of patients with HCC (4,9). This classification has been endorsed by an EASL panel of experts and the AASLD guidelines (7,8) and has been externally validated in European and American patient cohorts (10,11). The BCLC classification links stage stratification with a recommended treatment strategy and defines standard of care for each tumor stage. The less frequently used Tumor–Node–Metastasis (TNM)–American Joint Committee on Cancer classification is based on vascular invasion status as defined by pathological examination (12) and does not take into account the functional status of the liver, both of which are limitations for widespread clinical use. No established clinically applicable biological or genetic markers associated with clinical outcomes have been defined to classify patients with HCC.

The natural history and current therapeutic strategies for patients within different BCLC categories are shown in Figure 1 (57) and discussed below.

Early Stages

The natural history of early HCC (stage 0 and stage A) is unknown because most patients with early-stage HCC are treated with potentially curative therapies (resection, liver transplantation, or local ablation, either with radiofrequency [RF] or percutaneous ethanol injection [PEI]) that are associated with 5-year survival rates of 50%–70% (4,13). The best actual outcome reported in untreated patients with Child–Pugh class A disease and with single tumors is 20% survival at 5 years (14). No specific RCTs comparing

curative therapies with best supportive care have been conducted because it is assumed that treatment improves life expectancy. Direct comparisons between resection and local ablation have yielded inconsistent results, whereas comparisons of resection with liver transplantation are based only on cohort studies and case series, which are prone to treatment selection biases. Although positive results have been reported for adjuvant therapies such as retinoids, interferon, and other treatments in isolated RCTs, no adjuvant therapies are currently accepted as the standard of care for HCC patients who have undergone complete resection or local ablation. Furthermore, the role of neoadjuvant therapies in patients awaiting liver transplantation remains unclear (7). These are gray areas of research that represent an unmet clinical need.

Prognostic factors that have been identified in surgical case series of patients with early-stage HCC are size of the main nodule, multicentricity (single nodule ≤2 cm, single nodule 2–5 cm, two or three nodules ≤3 cm), and liver function, as assessed by portal hypertension and serum bilirubin (13). The Child–Pugh classification is used for prognosis estimation in patients undergoing percutaneous ablation. Patients with early tumors rarely present with cancer-related symptoms. In liver transplantation case series, variables related to early HCC status (single tumor ≤5 cm or three nodules ≤3cm)—the so-called Milan criteria—define good prognosis (15).

Intermediate to Advanced Hepatocellular Carcinoma

The prognosis of patients with unresectable HCC has been assumed to be poor, with median survival of less than 1 year. The 1- and 2-year survival rates of untreated patients randomly assigned to the control arm in 25 RCTs ranged from 10% to 72% and from 8% to 50%, respectively (5). These widely discrepant figures likely reflect the inclusion of patients with differing stages of the disease. Indeed, the BCLC classification further stratifies patients with unresectable HCC into three categories: intermediate (stage B), advanced (stage C), and end stage (stage D) (4,9). Untreated patients with intermediate-stage HCC (multinodular asymptomatic tumors without vascular invasion or extrahepatic spread) have a median survival of approximately 16 months (5). Chemoembolization improves median survival to 19–20 months in RCTs (5,16,17) and is considered the standard of care (7). Untreated patients with advanced-stage disease (ie, ECOG performance status 1 or 2 and/or vascular invasion or extrahepatic spread) have a median survival of 6–7 months. Among this group, outcomes vary according to Child–Pugh class. Recently, sorafenib—a multitarget tyrosine kinase inhibitor—was found to improve survival of patients with advanced-stage HCC (median overall survival was 10.7 months for sorafenib vs 7.9 months for placebo, P = .0006) (18). Sorafenib was approved by regulatory agencies during 2007 and is likely to become the new standard systemic therapy for HCC patients with advanced disease.

End-Stage Hepatocellular Carcinoma

Patients who present with end-stage disease (stage D) have high HCC tumor burden associated with an ECOG performance status of 3 or 4. These patients have a reported median survival of only 3 months (4). Similarly, Child–Pugh class C patients with HCC who are unsuitable for transplantation also have a very poor prognosis.

Evidence-Based Recommendations in Oncology

The levels of evidence for treatment recommendation in oncology provide a common framework to understand the evidence-based treatment recommendations in HCC. In general, management of cancer patients should be supported by high-quality evidence. Three levels of clinical decision making have been described (19). The first level is based on personal experiences and might be applied to small numbers of individuals. The second level is based on empirical experiences and is applied to hundreds of patients of a given physician. The third level represents across-the-board recommendations for a population—that is, it affects thousands of patients—and must be based on rigorous assessment of the scientific evidence. Treatment recommendations are based on the strength of the evidence and the magnitude of benefit (and the risk to benefit ratio).

Strength of Evidence

Clinical trials (phase 1, 2, and 3) are the mainstay of experimental studies and provide the most convincing evidence for any hypothesis. The criteria to define levels of evidence have been summarized by the US National Cancer Institute (20). A hierarchy is often used that incorporates both the strength of study design and the strength of endpoints (Table 1). In principle, the double-blinded RCT (category 1) is the gold standard of clinical trial design, but the RCT must be of high quality to reliably avoid confounding biases. Blinding is not mandatory when survival is the primary endpoint. Critical components of high-quality trial design include a clear definition of study population and endpoints, sample size and power calculations, treatment allocation and masking, stratification, intention-to-treat analysis, and interim analysis plans and stopping rules (21,22). These key components of RCT quality should be included in the assessment of clinical outcomes within phase 3 studies (see Table 4) (23). Guidelines for reporting RCTs are summarized in the CONSORT statements (24).

Nonrandomized studies (category 2, typically nonrandomized phase 2 studies) provide less robust evidence and are generally insufficient to change standards of care or warrant drug approval. Category 2 analyses also include subset analyses of RCTs, which are subject to errors inherent in multiplicity (ie, "statistically significant" results are expected as a result of random variation of measured effects in multiple subsets). Finally, case series and retrospective analyses provide the lowest strength of evidence (category 3).

Magnitude of the Benefit

Survival benefits derived from treatments in oncology are highly heterogeneous, and the definition of clinically meaningful survival benefits is controversial (25). On the basis of a cost-effectiveness perspective, some investigators have suggested that clinically relevant survival improvements are those exceeding 3 months (26). The magnitude of benefits may be relative to the outcome expected for the target population, either untreated or receiving the standard-of-care treatment, and the benefit/risk ratio. Therefore, the magnitude of benefit can be expressed as the relative percentage improvement in survival (or other, surrogate endpoint) or decrease in the hazard ratio of death.

Endpoints for Targeted Therapies in Oncology

The new targeted drugs present particular challenges to trial design and methodology in clinical research. Trials assessing traditional cytotoxic agents were designed to capture cytotoxic effects as a decrease in tumor cross-sectional area based on two-dimensional measurements (a surrogate for tumor volume, which until recently was difficult to measure). The assumption has been that a decrease in tumor size was a step toward achieving complete remission. Such tumor "shrinkage" (objective response) was also postulated to be clinically relevant and to translate to enhanced survival, a supposition that was often without concrete data. However, current targeted agents may act as cytostatic agents and/or increase inflammatory response and possibly improve survival with no measurable change in tumor size. Thus, for this class of agents time-to-event endpoints may be more critical than decreases in tumor volume.

In current oncological practice, phase 1 studies are intended to define appropriate dosage by using endpoints such as dose-limiting toxicity, maximum tolerated dose, pharmacokinetic profile, and pharmacodynamic profile. The primary endpoint of these studies is the safety profile or change in measures that reflect relevant biologic processes. Different trial designs and dose escalation strategies have been described elsewhere (21,22,27). Alternative trial designs for novel drugs could evaluate toxicity and estimate the inhibition of a target in surrogate cells, such as those found within the peripheral blood, and define a dose that is able to abrogate a specific signaling pathway whose alteration is expected to produce the desired biological effect in the tumor.

Phase 2 studies are typically designed to determine antitumor activity in a selected group of patients with a specific cancer. Assessment of molecularly targeted therapies may, however, require a time-to-event endpoint to properly capture this antitumor effect (28). Other endpoints that are generally assessed include toxicity, duration of response, and biomarker response. Phase 2 studies can be either single-arm studies or randomized studies with a control arm that includes the standard of treatment for the targeted population. Randomized phase 2 studies have been used extensively in oncology in recent years. Of 226 randomized phase 2 studies analyzed in a recent review (29), the median sample size was 40 patients per arm, and only about 15% of drugs were advanced to phase 3 studies.

Phase 3 studies are aimed at assessing clinically relevant outcomes. These studies are needed to provide a sufficient level of evidence of survival advantage or other clinical benefit for any treatment to change clinical practice. The eligibility criteria determine the study population but are a potential source of selection bias that can affect generalizability of the results. Populations enriched by specific biomarkers in so-called targeted trials may permit a large reduction in the number of patients that are needed, but the results will not be readily generalizable due to the lack of reliable biomarkers and unknown off-target effects.

The most common endpoints assessed in oncology trials are overall survival, disease- and progression-free survival, time to progression, and response rate. Determination of all endpoints other than survival can be subjective. The endpoints that were used the most by the US Food and Drug Administration (FDA) for

Table 1. Levels of evidence in the assessment of benefits in the treatment of hepatocellular carcinoma according to the strength of study design and of endpoints*

Treatments assessed	Benefit	Evidence
Surgical treatments		
Surgical resection	Increased survival	3iiA
Adjuvant therapies	Uncertain	1iiA
Liver transplantation	Increased survival	3iiA
Neoadjuvant therapies	Treatment response	2iiDiii
Locoregional treatments		
Percutaneous treatments	Increased survival	3iiA
Percutaneous ethanol injection		
Radiofrequency ablation	Better local control	1iiD
Chemoembolization	Increased survival	1iiA
Arterial chemotherapy	Treatment response	3iiDiii
Internal radiation (I131, Y90)	Treatment response	3iiDiii
Systemic treatments		
Sorafenib	Increased survival	1iA
Tamoxifen	No benefit	1iA
Systemic chemotherapy	No benefit	1iiA
Interferon	No benefit	1iiA

* Classification of evidence adapted from National Cancer Institute: www.cancer.gov. Twenty classifications are as follows: Study design: randomized controlled trial, meta-analysis = 1 (double-blinded: 1i, nonblinded: 1ii). Nonrandomized controlled trials = 2. Case series = 3 (population-based 3i, non–population-based, consecutive 3ii, non–population-based, nonconsecutive: 3iii); endpoint: survival (A), cause-specific mortality (B), quality of life (C). Indirect surrogates (D) (disease-free survival [Di], progression-free survival [Dii], tumor response [Diii]).

regular and accelerated approval of anti-neoplastic drugs during 1990–2002 were overall survival and antitumor response (30).

Design of Clinical Trials in Hepatocellular Carcinoma: Endpoints

Clinical trials in HCC should be designed in accord with conventional biostatistical rules applied in oncology trials (21,22,27). When selecting endpoints in HCC clinical trials, researchers should give specific consideration to the fact that death often occurs as a result of liver failure in patients with liver cancer (Box 1). In principle, there are two general categories of endpoints in clinical trials: direct clinical outcomes (which include overall survival, time to recurrence, and time to symptomatic progression [TTSP]) and potential indirect surrogates.

Direct Clinical Outcomes

Overall Survival. This outcome captures the time from random assignment until death. It is the most important endpoint and the one that is least subject to investigator bias. The FDA relied on improved survival as an endpoint for regular approval of drugs in oncology in 18 of 57 cases (31%) from 1990 to 2002 (30). Overall survival was the primary endpoint recommended by the expert panel for any phase 3 study in HCC. Although this outcome should also be reported in phase 2 studies, it does not serve as a reliable endpoint for this study design because it cannot be properly powered.

Cancer-specific survival is a related endpoint, in which only deaths due to cancer are considered for survival analysis and non–cancer-related deaths are censored. Although this endpoint may be of biologic importance in a disease-specific intervention, it is a

> **Box 1.** Endpoints in clinical trials in hepatocellular carcinoma.
>
> **Recommended primary and secondary endpoints**
>
> **Survival:** Time from randomization to death. Patients alive at the end of follow-up are censored.
> - Primary endpoint in phase 3 studies assessing primary treatments.
> - Primary/secondary endpoint in phase 2/3 studies assessing adjuvant or neoadjuvant treatments.
> - Secondary endpoint in phase 2 studies assessing primary treatments.
>
> **Time to recurrence (TTR)*:** Time from randomization to recurrence. Evidence of recurrence should follow the Response Evaluation Criteria in Solid Tumors (RECIST) amendments (see text for details). Once evidence of hepatocellular carcinoma (HCC) recurrence is confirmed, TTR will be defined as the time that recurrence was first suspected.
> - Primary/secondary endpoint in phase 2/3 studies assessing adjuvant or neoadjuvant treatments.
>
> **Time to progression*:** Time from randomization to radiological progression. Definition of progression is based on the RECIST amendments. Deaths during follow-up without evidence of radiological progression are censored.
> - Primary endpoint in phase 2 studies assessing primary treatments.
> - Secondary endpoint in phase 3 studies assessing primary treatments.
>
> **Time to local recurrence*:** Time from randomization to local radiological progression. Definition of progression is based on the RECIST amendments. Deaths during follow-up without evidence of radiological progression are censored.
> - Secondary endpoint in studies assessing locoregional therapies.
>
> **Tertiary endpoints†**
>
> **Cancer-specific death:** Time from randomization to HCC-related death. Patients alive at the end of follow-up are censored.
> - Competing risk analysis is recommended to assess this endpoint.
>
> **Time to symptomatic progression:** Time from randomization to deterioration of symptoms as assessed by a standardized questionnaire.
> - No reliable questionnaires have been thoroughly validated in HCC research.
>
> **Disease-free survival:** Composite endpoint. Time from randomization to either recurrence or death. Patients alive and free of recurrence at the end of follow-up are censored.
> - Vulnerable endpoint in HCC research.
>
> **Progression-free survival:** Composite endpoint. Time from randomization to either radiological progression or death. Patients alive and free of progression at the end of follow-up are censored.
> - Vulnerable endpoint in HCC research.
>
> **Response rate:** Definition of response is based on the RECIST amendments.
>
> ---
>
> * Time to progression and time to local recurrence can vary considerably if evaluation interval varies among studies or between study arms of an individual study.
>
> † Tertiary endpoints include composite endpoints that are vulnerable in HCC research, such as disease-free and progression-free survival, that are difficult to measure with standard tools, such as time to symptomatic progression, or that are not time-to-event endpoints, such as response rate or disease control rate.

more subjective endpoint than total mortality. In particular, whereas death in Child–Pugh class A patients with HCC is usually related to tumor progression, it can be difficult to accurately ascertain the cause of death in HCC patients with Child–Pugh B or C disease. Treatment-related deaths may also be missed in the analysis of cancer-specific survival.

The assessment of cancer-specific survival needs to take into account intercurrent mortality (ie, mortality from causes other than HCC) because even modest intercurrent death rates can dilute the evaluation of treatment differences (31). The standard method to estimate the probability of death over time is Kaplan–Meier statistical analysis. However, this method does not allow for the recognition of other clinical outcomes, including death from other causes (eg, liver failure), and can thus overestimate the probability of death from HCC. In contrast,

the competing risk analysis method will provide an estimate of HCC-related death in the presence of alternate yet plausible outcomes such as death from liver failure or liver transplantation. Notably, this approach has been used successfully to highlight differential rates of competing outcomes in patients waiting for a liver transplantation (32).

The following example illustrates the relevance of accounting for competing risks depending upon the endpoint chosen. If an intervention is hoped to reduce mortality (ie, the endpoint is overall survival) from 30% to 15%, the risk ratio is 2, with a binomial sample size calculation of 320. However, if the intervention is hoped to reduce HCC-related death (ie, the endpoint is cancer-related death), and the competing mortality from progressive liver failure is 15% in both arms, the risk ratio will be reduced to 1.5 and the sample size consequently increased to 440. Thus, a competing risk

analysis that assesses cancer-related deaths requires a larger sample size than an overall survival analysis. A similar competing risk approach should be considered when assessing time to recurrence.

Time to Recurrence. Time to recurrence was recommended by the panel as the primary endpoint for HCC phase 2 and 3 studies that assess adjuvant therapies after resection or local ablation. This endpoint is more difficult to interpret in single-arm phase 2 studies than in randomized phase 2 trials because of the lack of a control group. Molecular studies have shown that recurrence after resection has two components. The main component, which accounts for 60%–70% of recurrences, includes true metastasis that results from HCC dissemination before resection and is undetectable by imaging techniques (13). This type of recurrence occurs mainly within the first 2 years after resection (33). The other component includes metachronous tumors that arise de novo in a preneoplastic cirrhotic liver. The type of tumor recurrence should be confirmed by molecular studies (eg, comparative genomic hybridization analysis, microarray analysis) (34,35) if feasible because treatments that are effective against metastasis may not prevent de novo cancer, and vice versa. Thus, molecular studies that aim to differentiate the two types of recurrences within RCTs are recommended, although they might be pragmatically difficult to accomplish.

Time to Symptomatic Progression. This endpoint reflects the time between random assignment and the occurrence of disease-related symptoms or differences in preestablished questionnaire scores. Time to symptomatic progression and health-related quality-of-life instruments have not generally been used as a sole basis for drug approvals. In general, time to symptomatic progression can capture deterioration in quality of life along with drug-related toxicity. This endpoint, however, is particularly hard to measure in cirrhotic patients with cancer, in whom the impairment of quality of life may be a consequence of the natural history of cirrhosis and not of tumor progression. In fact, there is no validated tool or questionnaire to measure quality of life in HCC. The most frequently used instrument is the Functional Assessment of Cancer Therapy Hep-30 scale (36), which was developed for patients with hepatobiliary tumors, and the European Organisation for Research and Treatment of Cancer QCL-C30 (37). The lack of any clear evidence that these are reliable questionnaires to capture this endpoint in HCC prevents their general recommendation for research purposes. Thus, the panel concluded that the time to symptomatic progression endpoint is not ready for HCC clinical research at this point. Nevertheless, the exploration of clinically relevant measures of changes in physical status and quality of life merits further investigation.

Potential Indirect Surrogate Endpoints

Surrogate endpoints are all usually subject to investigator interpretation. More important, they do not automatically translate into direct patient benefit. For example, in RCTs of paclitaxel in patients with metastatic breast cancer, improvements in response rate and progression-free survival were not followed by improvements in health outcomes (38). Nevertheless, in many cases regulatory agencies have approved treatments that improve these surrogate endpoints while awaiting a more definitive endpoint to support their use. In this regard, tumor response rate, either alone or together with time to progression, was the endpoint used by FDA to justify regular approval of 26 drugs in oncology (46%) or accelerated approval of 12 drugs (85%) between 1990 and 2002 (30).

Disease-Free Survival and Progression-Free Survival. These are composite endpoints that include two types of variables: death and evidence of radiological recurrence or death and evidence of radiological progression (disease- and progression-free survival, respectively). In general, regulatory agencies prefer progression-free survival to time to progression for drug approval because the former endpoint may be better correlated with overall survival. However, although disease- and progression-free survival are appropriate endpoints in other solid tumors, they are particularly unreliable endpoints in HCC research because death resulting from the natural history of cirrhosis might confound detection of potential benefits from effective drugs. That is, a type II error might result from using progression-free survival as an endpoint in a suboptimal population in early phases of drug development. In the scenarios shown in Figure 2, benefits in preventing progression could be masked by death as a result of liver failure (Figure 2, A), whereas two drugs that actually have the same effects on progression might appear to differ in their benefits (type I error) because of imbalances in liver failure–related death (Figure 2, B). Thus, the panel discouraged the use of progression-free survival as a primary endpoint for trials testing new compounds in HCC. Similarly, disease-free survival is not supported for assessment of adjuvant therapies after resection, transplantation, or local ablation because of the confounding composite nature of this endpoint. In the exceptional circumstances in which these endpoints are applied, a restrictive selection of patients with well-preserved liver function is recommended to minimize the impact of death unrelated with tumor progression.

Time to Progression. This endpoint reflects the time between random assignment and radiological progression as defined by the amendments of the Response Evaluation Criteria in Solid Tumors (RECIST; 39), as described below ("Radiological measurements of Time to Progression and Response Rate"). This pure radiological endpoint requires repeated radiological measurements to capture relevant differences between groups that can be missed if the intervals between measurements are too long. The formal recommendation of the panel was to conduct imaging surveillance every 6–8 weeks by computed tomography (CT) scan or magnetic resonance imaging (MRI). Symmetric assessment should be ensured between treatment arms. Deaths are censored as nonradiological progressions at the time of death or at an earlier visit, representing informative censoring. The panel recommended time to progression as the main time-to-event endpoint to capture possible antitumor benefits in phase 2 trials testing molecular targeted therapies in HCC because it is less vulnerable (only progression captured) than composite endpoints.

Robust evidence is needed to support time to progression as a true surrogate of overall survival in phase 3 trials of HCC. However, to date, this endpoint has rarely been measured in phase 3 studies of HCC. Formal application of criteria for true surrogacy of time to progression for overall survival in HCC is awaited from the analysis of the phase 3 trial of sorafenib (18).

Figure 2. Time to progression (TTP) and progression-free survival (PFS) assessment in hepatocellular carcinoma (HCC). The ability of these endpoints to capture treatment benefit is illustrated for two possibilities (treatment A [**dashed lines**] is better than treatment B [**solid lines**] or treatment A is the same as treatment B) under two trial design scenarios (a theoretical trial with good design [ie, no potential bias included—inclusion criteria are correct, there is no treatment toxicity, and all deaths are from HCC] and a real scenario [ie, potential bias included—inclusion criteria may not be correct, there may be treatment toxicity, and some deaths may be due to cirrhosis or treatment toxicity instead of to HCC]). The graphs plot the percentage of patients free of progression (PFS) against time in months and the cumulative incidence of progression (TTP) against time. PFS is calculated as $1 - TTP$. **A**) Assessment of TTP and PFS when drug A is better than drug B. In the theoretical scenario (**left panel**), both endpoints reveal a statistically significant difference between the arms. In the real scenario (**right panel**), TTP continues to show a statistically significant difference between treatments. However, because of confounding deaths as a result of liver failure, PFS is unable to capture the true benefits from drug A, leading to a type II error. **B**) Assessment of TTP and PFS when the effects of drug A are not different from those of drug B. In the theoretical scenario, both endpoints show non–statistically significant differences between the two arms. However, in the real scenario, confounding deaths can result in PFS indicating an antitumor effect when no such effect exists (type I error). **C**) Assessment of TTP when drug A is better than drug B according to Kaplan–Meier analysis (**left panel**) or a competing risks assessment (**right panel**) under the real scenario.
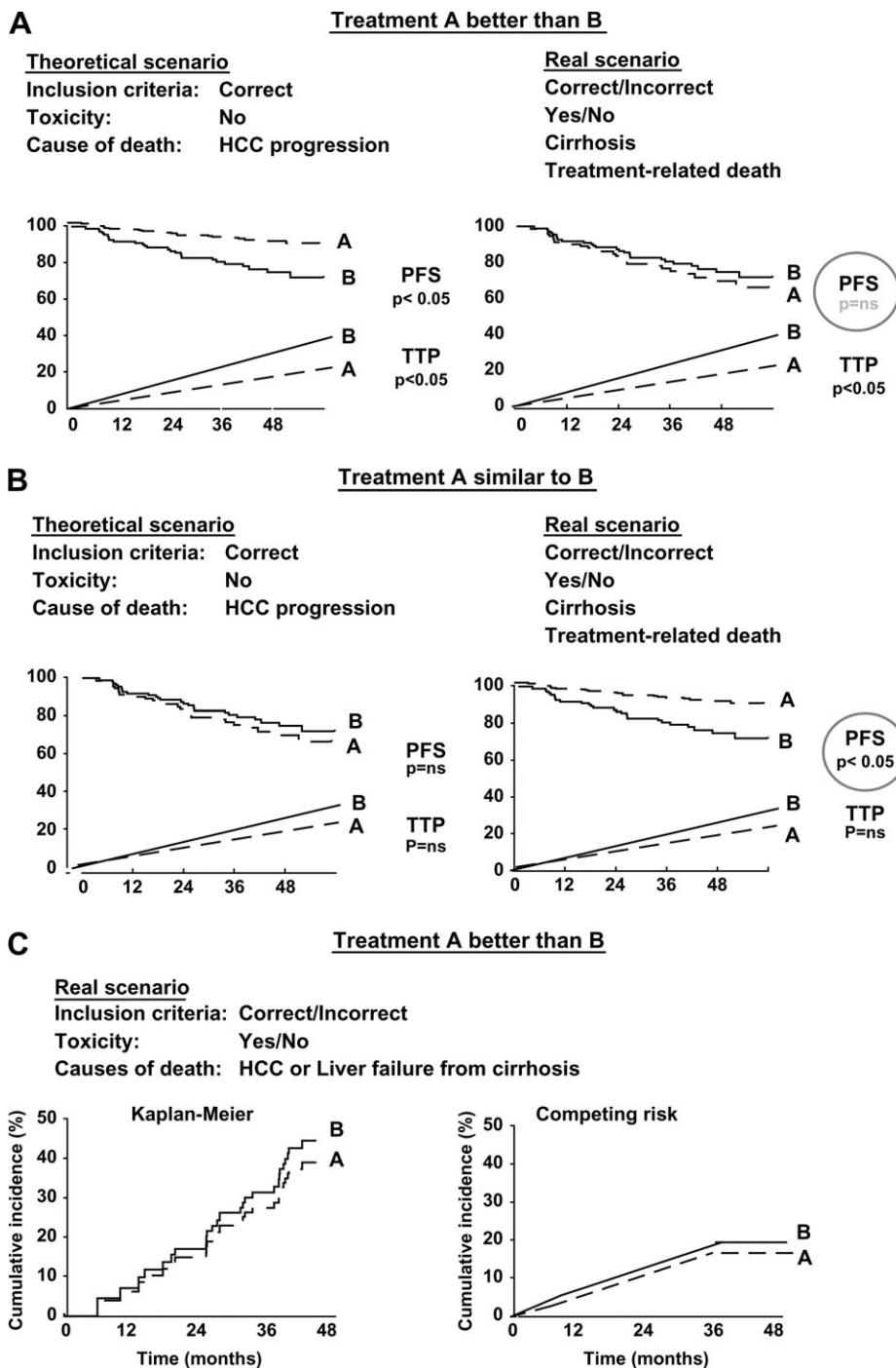
Response Rate. Antitumor response, as measured according to the World Health Organization (39) and RECIST (40) criteria, has been considered to be the primary endpoint for phase 2 studies to proceed to further phase 3 investigations. Studies applying Cox proportional hazards analysis in HCC research suggest that this endpoint is consistently associated with survival in trials of locoregional therapies (17). Tumor shrinkage resulting from necrosis, ischemia, and cytotoxicity induced by RF ablation and chemoembolization precedes a survival benefit, fulfilling one of the requirements for true surrogacy. However, with the advent of targeted molecular compounds the reliance on response rate

needs to be reconsidered because clinically significant survival advantages have been reported with only marginal tumor responses. Indeed, molecularly targeted compounds that produce objective response rates of less than 10% have resulted in improved survival in RCTs, including erlotinib in non–small cell lung cancer (41), temsirolimus in renal cancer (42), bevacizumab in metastatic colorectal cancer (43), and, more recently, sorafenib in liver cancer (18) (Table 2). Thus, the panel formally discouraged the use of response rate as an endpoint for capturing the benefits of targeted drugs in phase 2 studies of HCC. This represents a change in the paradigm of design of clinical trials in

Table 2. Change in the paradigm for designing clinical trials in hepatocellular carcinoma

Type of treatment	Objective response	Survival benefit
Conventional treatments in HCC		
Local ablative therapies (RF ablation and/or PEI)	70–80% (CR)	Yes
Chemoembolization	35–40% (PR)	Yes
Internal radiation (I131, Y90)	20–30% (PR)	Unknown
Intraarterial chemotherapy	15–20% (PR)	Unknown
Systemic chemotherapy	~10% (PR)	No
Molecular targeted therapies in oncological practice†		
Small-molecule kinase inhibitors		
EGFR: erlotinib (NSCLC) (41)	9% (PR)	Yes
Raf/VEGFR: sorafenib (HCC) (18)	2.7% (PR)	Yes
mTOR: temsirolimus (RCC) (42)	8% (PR)	Yes
Monoclonal antibodies		
Anti-VEGF: bevacizumab (metastatic CRC) (43)	10% (PR)	Yes

* HCC = hepatocellular carcinoma; RF = radiofrequency; PEI = percutaneous ethanol injection; CR = complete response; PR = partial response; NSCLC = non–small cell lung cancer; RCC = renal cell carcinoma; CRC = colorectal cancer; VEGF = vascular endothelial growth factor; EGFR = epidermal growth factor receptor; VEGF = vascular endothelial cell growth factor; mTOR = mammalian target of rapamycin.

† Examples of targeted therapies that improve survival with an objective response rate of less than 10%. Reference numbers are in parentheses.

oncological practice in general and in the study of new treatments for HCC in particular (44).

Radiological Measurements of Time to Progression and Response Rate

The expert panel recommended adopting a modification of the RECIST criteria (39) to assess response rates and time to progression. These endpoints should be defined following centralized radiological review rather than being based on the assessment of local investigators.

The RECIST (39) criteria define standard methods for converting radiology image observations to a quantitative and statistically tractable framework for measuring the response of tumor size to therapy. Target lesions are measured using a single linear summation. These criteria were designed primarily for evaluation of cytotoxic agents that induce cell death, even in the absence of major necrosis. Hence, assessments based solely on changes in tumor size or morphology can be misleading when applied to other anticancer drugs, such as molecular targeted therapies, or other therapeutic interventions. Recent studies have found a poor correlation between the extent of tumor necrosis induced by new agents such as sorafenib or by interventional procedures such as chemoembolization and conventional methods of response assessment. The original RECIST publication (39) did not address measures of antitumor activity other than tumor shrinkage. In 2000, however, a panel of experts convened by EASL recommended that the response criteria be amended to take into account tumor necrosis induced by treatment (8). That panel considered estimation of the reduction in viable tumor area using contrast-enhanced radiological imaging to be the optimal method to assess treatment response. Viable tumor was defined as uptake of contrast agent in the arterial phase of dynamic CT or MRI.

The expert panel has adapted the concept of viable tumor endorsed by EASL (7) and AASLD (8) and proposed the following amendments to RECIST criteria in the determination of tumor response for target lesions in HCC: complete response is the disappearance of any intratumoral arterial enhancement in all target lesions; partial response is at least a 30% decrease in the sum of diameters of viable (contrast enhancement in the arterial phase) target lesions, taking as reference the baseline sum of the diameters of target lesions; progressive disease is 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 the treatment started; stable disease is any cases that do not qualify for either partial response or progressive disease.

According to these amendments, progressive disease will also be declared on the appearance of one or more new lesions, as per RECIST (39). A newly detected hepatic nodule will be classified as HCC—and therefore will be declared as evidence of progression—when its longest diameter is at least 10 mm and the nodule shows the typical vascular pattern of HCC on dynamic imaging, that is, hypervascularization in the arterial phase with washout in the portal venous or late venous phase. Lesions larger than 10 mm that do not show a typical vascular pattern can be diagnosed as HCC by evidence of at least 1-cm interval growth in subsequent scans. An individual radiological event will be adjudicated in retrospect as progression at the time it was first detected by imaging techniques, even if strict criteria were fulfilled only on subsequent radiological testing. Finally, for nonenhancing atypical lesions, the conventional RECIST criteria will be applied.

These definitions also apply to the assessment of HCC recurrence after resection or local ablation. As per RECIST, cytopathological confirmation of the neoplastic nature of any effusion (particularly ascites) that appears or worsens during treatment is required when the measurable tumor has met criteria for response or stable disease.

To properly use the proposed criteria to improve comparability across studies, uniform image acquisition parameters, quality control, and independent blinded multireader assessments should be used. In view of the accuracy of volumetric determinations that are performed by current imaging instruments, we suggest that direct volumetric measurement to identify partial response and progression should be a priority in future clinical trial research.

New Tools in Hepatocellular Carcinoma Clinical Research

Molecular Imaging. The FDA has established the "Critical Path Initiative," whose goal is to facilitate innovation in drug discovery and clinical trial development; functional imaging is a major cornerstone to this initiative (45). A number of quantitative molecular imaging techniques can be used to study tumor physiology in both the preclinical and clinical arenas. For example, in various cancers the pattern of tumor microvascular perfusion with dynamic contrast-enhanced MRI has been shown to correlate with histological grade, microvessel density, vascular endothelial growth factor expression level, and the effects of antiangiogenic and antivascular therapy (46).

Further studies are required to confirm these relationships in HCC. Currently, no molecular imaging method has been shown to accurately detect, characterize, or monitor the response of HCC to treatment. Difficulties in developing specific imaging methods for HCC are caused by the lack of obvious specific molecular targets, problems with drug delivery, and poor signal-to-noise ratios (47). Nevertheless, advances in molecular imaging are expected to help identify patients who are most likely to benefit from intervention, particularly with molecular targeted drugs, and to provide new parameters of interpreting efficacy.

New Biomarkers. Few biomarkers have been used in clinical research in oncology, and none have been proven to be an adequate surrogate for true health outcomes such as overall survival. With only a few exceptions, including paraprotein levels in blood and urine, which have been used as part of myeloma response criteria (48), and germ cell tumor markers, biomarkers have not served as primary endpoints for cancer drug approval. Alpha-fetoprotein (AFP) has been suggested as a biomarker for measuring antitumor response or progression in HCC, but there are insufficient data to suggest its use in clinical research. Moreover, fluctuations of AFP levels can result from flares of viral reactivation that are unrelated to cancer development. A decrease in AFP levels is associated with clear tumor regression in less than one-third of HCC patients (49). Further research is therefore required to assess the value of this marker in clinical trials of HCC.

Design of Clinical Trials in Hepatocellular Carcinoma: Study Population, Stratification, and Control Arms

Study Population

Diagnosis of Hepatocellular Carcinoma. Pathological confirmation of HCC or noninvasive criteria following the AASLD guidelines (7)—recently validated (50)—were accepted by the expert panel for trial eligibility. In brief, HCC can be defined in cirrhotic patients by one imaging technique (CT scan, MRI, or second-generation contrast ultrasound) showing a nodule larger than 2 cm with contrast uptake in the arterial phase and washout in venous or late phases or two imaging techniques showing this radiological behavior for nodules of 1–2 cm in diameter. Cytohistological confirmation is required for patients who do not fulfill these eligibility criteria. Molecular diagnosis of HCC is not yet ready for clinical research (51).

Target Population. It is important to select homogeneous patient populations for the evaluation of new agents. Otherwise, the results of clinical trials will be difficult to interpret and can be confounded by differing patient characteristics. The BCLC classification defines homogeneous populations of patients with different natural outcomes and specific prognostic variables (4,9). Thus, patients included in an HCC clinical trial should either represent a specific BCLC class or be stratified for this staging system. Although patients with both Child–Pugh class A and B disease can be considered for clinical trials in HCC research, the working group recommended an initial focus on Child–Pugh class A patients. These selection criteria facilitate the assessment of drug effect without the confounding issues of liver failure and death as a result of underlying cirrhosis. A recent systematic review (52) of 118 studies of patients with cirrhosis but no known HCC found that Child–Pugh class A patients have 1- and 2-year actuarial survival rates of 95% and 90%, respectively, compared with 80% and 70%, respectively, in Child–Pugh class B patients and 45% and 38%, respectively, in Child–Pugh class C patients. Accordingly, in Child–Pugh class B or C patients, death from cirrhosis could potentially mask treatment-related antitumor efficacy. After initial studies of a new agent or combination of agents in Child–Pugh A patients, subsequent studies can be conducted in HCC patients with Child–Pugh B disease to assess the safety profile of new compounds in this population. Child–Pugh class C patients are poor candidates for any clinical research because of their poor prognosis without liver transplantation.

Parameters Assessed and Reported in Clinical Trials. The minimum baseline variables that should be included in HCC phase 2 and 3 studies are detailed in Table 3. Specific surgical findings should be reported in adjuvant studies. Race/ethnicity and sex of patients should be reported in all trials because the natural history of HCC and treatment effect may vary across these categories.

Stratification Before Randomization

The expert panel recommended that the target population for an experimental study in HCC should be selected according to the BCLC staging system (Figure 1). The target population should also be stratified by BCLC stage before randomization; post hoc statistical adjustment is less desirable as a means to control for unbalanced prognostic variables. Among patients with advanced-stage HCC (BCLC stage C), stratification for ECOG performance status (0 vs 1–2), tumor burden (vascular invasion and/or extrahepatic spread), and Child–Pugh class are recommended. Stratification by region of the world is advisable in international trials (Western countries vs Asian countries). "Overstratification" by less important prognostic variables such as etiology and age is not desirable; however, randomization can provide balanced distributions of these variables. Stratification using other staging systems, such as the Cancer of the Liver Italian Program (CLIP) (53) or Chinese University Prognostic Index (CUPI) (54), was generally discouraged by the panel because they have suboptimal capacity to define homogeneous subgroups among patients with advanced- and end-stage disease.

Trials of adjuvant therapy after potentially curative therapies should be generally conducted in patients of BCLC stage 0 or A. In studies that have tissue specimens available, patients should be stratified for high or low risk of recurrence. High risk of recurrence after resection is defined by the presence of vascular invasion, satellite nodules/multinodularity, or poor differentiation (13). Pathological TNM stage information should also be collected in such trials, although the prognostic implications of this classification are less clear. In the case of adjuvant trials after ablation, patients should be stratified by tumor size and multinodularity, which reflect the risk of recurrence (55,56).

Control Arm of Randomized Trials

Patients randomly assigned to the control arm in a clinical trial should receive the best standard of care. In 2008, both PEI and RF ablation are considered the standard of care for patients with early HCC, with no proven survival difference between them (6,7). For comparisons of new treatments in patients with intermediate-stage (BCLC stage B) disease, chemoembolization is the standard of care, based on RCTs (5,16,17). Until recently, new agents for patients with more advanced HCC (BCLC stage C) have been mostly compared with either placebo or best supportive care. The demonstration of improved survival using the multitarget kinase inhibitor sorafenib in a placebo-controlled trial has changed this situation (18). Most investigators are now likely to accept sorafenib as the standard of care, and the expert panel cautiously recommended it as a control arm for future RCTs of first-line systemic agents. Ongoing trials should seek the advice from their Data Safety and Monitoring Committee (DSMC) anytime a new drug or device demonstrates clinical benefits for the same HCC target population.

Liver Function in Trial Design

Cirrhotic patients present specific challenges to management and interpretation of treatment toxicity in trials of new agents. Such patients are prone to liver decompensation from underlying cirrhosis or reactivation of chronic hepatitis B viral infection after local or systemic chemotherapy. Therefore, whenever possible, studies should separately include (and/or analyze) patients with and without cirrhosis. The definition of cirrhosis should be identified and reported as either biopsy proven or based on clinical and radiological features. Although Child–Pugh score was endorsed by the panel as the gold standard for measurement of liver impairment, the panel also recommended that future trial also report the Model of End-Stage Liver Disease score (57). Once a clinical benefit of any therapy has been identified in Child–Pugh class A patients, further studies should be designed to confirm the efficacy in patients with more advanced liver impairment.

Stopping rules for reasons of safety and toxicity should be defined by the DSMC. Liver-related toxic effects are captured by serum aminotransferase levels, bilirubin levels, and prothrombin time. Liver decompensation and liver failure–only deaths regardless of treatment-associated antitumor effects are expected. The actuarial probability of 1- and 2-year deaths due to liver failure regardless of treatment response is 5% and 10% in Child–Pugh A class patients (52). Treatment-related deaths should be less than 3% for all the interventions.

Table 3. Variables to be included in clinical trials assessing treatments for hepatocellular carcinoma patients*

Demographic	Age, sex, ethnicity
	Underlying liver disease (cirrhosis, chronic hepatitis)
	Etiology (HCV, HBV, alcohol, hemochromatosis, others)
Tumor description	Radiological characteristics: size, number of nodules, macroscopic vascular invasion, extrahepatic spread
	Alpha-feto protein
	Pathological characteristics in adjuvant trials: size, differentiation degree, satellites, micro- and macroscopic vascular invasion, pTNM
Staging system	BCLC staging classification
Liver function	Bilirubin, aminotransferases, albumin, alkaline phosphatase, gamma-glutamyl transpeptidase, BUN, serum creatinine, serum sodium, prothrombin time, platelet count
	Presence of ascites or encephalopathy
	Child–Pugh class
	MELD score
General health	Performance status according to ECOG, pain, constitutional syndrome

* Modified from the European Association for the Study of the Liver consensus conference (7). Other variables might be recommended depending on the purpose of the trial, such as specific biomarkers or histological parameters (necrosis, inflammatory infiltrate). HCV = hepatitis C virus; HBV = hepatitis B virus; pTNM = pathological tumor–node–metastasis; BCLC = Barcelona Clinic Liver Cancer; BUN = serum urea nitrogen; MELD: Model of End-Stage Liver Disease; ECOG = Eastern Cooperative Oncology Group.

Summary of Trial Design Strategies in Hepatocellular Carcinoma

The strategy recommended by the expert panel for trial design in HCC research is summarized in Figure 3 and Table 4. These recommendations are intended to provide a framework as of 2008 that will evolve as new evidence becomes available, including more precise information on the natural history of HCC, novel therapies, and predictive biomarkers.

Design of Phase 1 Studies

Specific phase 1 studies in Child–Pugh class A patients with cirrhosis and HCC are recommended to estimate the exact dose, toxicity, and liver-related event risk that are not captured by phase 1 studies that include patients with a variety of neoplasms.

Single-arm proof-of-concept phase 2 studies are recommended as part of phase 1/2 studies, in which patients in the phase 2 part of the trial will be tested at the adequate dose.

Design of Phase 2 Studies

Randomized phase 2 studies were proposed by Simon et al. (58) to minimize the likelihood of erroneous conclusions regarding efficacy. This type of design is recommended by the expert panel for testing new drugs and devices in HCC research. Consequently, single-arm phase 2 trials are generally discouraged. However, large single-arm phase 2 studies might be considered when a contemporary historical control arm has been well characterized within other trials, and,
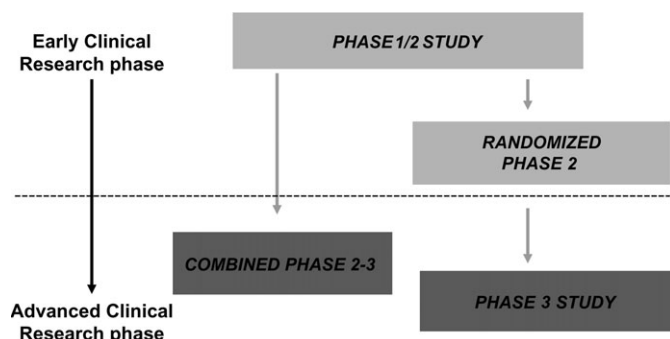
Figure 3. Hepatocellular carcinoma (HCC) trial design strategy recommended by the expert panel. Combined phase 1/2 studies in cirrhotic patients will capture liver-specific toxic effects of new drugs. Randomized phase 2 trials are pivotal in HCC research because they enable reliable comparison with the standard of care. Phase 3 studies remain the main source of evidence in HCC research. Phase 2/3 studies may be appropriate in very specific circumstances, that is, where the drug proposed would require fast-track assessment.

thus, inclusion criteria can be reproduced. In these exceptional circumstances, the results of single-arm phase 2 studies in patients with advanced HCC can be compared with a well-known outcome in a homogeneous control group that used the same inclusion criteria. Any such comparison should still be treated with caution.

The panel recommended that randomized phase 2 studies be powered (for instance, type I error 10%, power 80%) to test a time-to-event surrogate endpoint such as time to progression to aid in the decision to advance to phase 3 studies. Overall survival should be reported as a secondary endpoint, along with a safety assessment. Response rate and disease control rates are recommended only as ancillary information. Composite endpoints such as disease- and progression-free survival should be tested only as secondary or tertiary endpoints. Quality-of-life assessment in HCC research is recommended only as ancillary information. Additional research in the area of quality of life will be important. Randomized studies that test molecular targeted therapies should optimally include biomarker analysis (in tissue and/or serum samples) to enable the identification of molecular markers of response and for pharmacokinetic purposes, as reported in other cancers (59,60).

The control arm for initial therapy of advanced HCC should be sorafenib. New agents that are being assessed as second-line treatments for advanced HCC should be compared with placebo or best supportive care. Only agents that have shown clinically meaningful benefit as second-line treatments in phase 3 investigations should be considered for a direct comparison with the approved first-line option in phase 2 studies. Similarly, new compounds being tested for use as adjuvant or neoadjuvant treatments should be compared with placebo or best supportive care in phase 2 studies.

Table 4. Conventional and proposed trial design, study population, and endpoints in hepatocellular carcinoma research*

Trial phase and component	Conventional design	New proposed design
Phase 1		
Study population	All cancers	HCC by BCLC Child–Pugh A
Study design	Phase 1	Phase 1/2
Aim	Dose defining	Dose defining
	Safety	Safety
Endpoint	Toxic effects	Toxic effects
	MTD, pharmacokinetics	MTD and/or OBD
	Dose defining	
Phase 2*		
Study population	Unresectable HCC	HCC by BCLC
	Child–Pugh A and B	Child–Pugh A
Study design	Single arm	Randomized phase 2
		Single arm†
Aim	Antitumor activity	Antitumor activity
	Safety	Safety
Endpoint	Response rate	TTP
	Toxicity	Survival, toxicity
Phase 3†		
Study population	Unresectable HCC	HCC by BCLC
	Child–Pugh A and B	Child–Pugh A
Study design‡	RCT	RCT
		Combined phase 2/3
Aim	Clinical outcome	Clinical outcome
Endpoint	Survival	Survival
	Response	TTP
	PFS, DFS	Recurrence§

* HCC = hepatocellular carcinoma; BCLC = Barcelona Clinic Liver Cancer; MTD = maximum tolerated dose; OBD = optimal biological dose; TTP = time to progression; RCT = randomized controlled trial; PFS = progression-free survival; DFS = disease-free survival.

† Large single-arm phase 2 studies might only be considered when a contemporary historical control arm has been well characterized within other trials, and thus, inclusion criteria can be reproduced.

‡ Consider phase 2/3 studies for fast-track approval with strong interim analysis.

§ Time to recurrence as primary endpoint in adjuvant trials.

Design of Phase 3 Studies

The primary endpoint for phase 3 studies that assess primary HCC treatments is survival. Endpoints such as time to progression must be validated as surrogate endpoints of survival before their routine incorporation as primary endpoints; at the present time they should be considered only secondary endpoints. Composite endpoints should be avoided in phase 3 studies in HCC research. Clinical trials of locoregional therapies should report a time-to-local recurrence endpoint. Phase 3 studies of adjuvant therapies should use either overall survival or time to recurrence as a primary endpoint. Clinical trials of treatment for patients awaiting liver transplantation should consider time to transplantation (or time to dropout) as the key primary endpoint.

High-quality phase 3 studies in HCC should include clear definition of interventions, initial assembly of comparable groups (stratification of patients before randomization according to the BCLC staging system or other established prognostic factors in adjuvant studies), adequate concealment (computer-generated allocation sequence, centralized randomization), double-blinded studies to warrant masking of outcome assessment when feasible, and an adequate relevant outcome endpoint (Supplementary Table 4, available online).

The control arm for phase 3 studies should be the standard-of-care therapy. Therefore, for the assessment of first-line systemic treatments for advanced HCC, the recommended trial design is one that would compare the combination of a new agent plus sorafenib with sorafenib alone. Comparison of single agents head to head with the standard-of-care therapy might jeopardize study approval and recruitment of patients due to ethical reasons, unless the tested agent showed promising efficacy in early phase 2 studies. For second-line treatments, the new agent should be tested against placebo or best supportive care, and the selection criteria should include patients with contraindications to or failures on sorafenib.

Dual phase 2/3 studies can overcome the inherent time limitations in study design and/or execution of two consecutive studies. The primary advantages of phase 2/3 studies are both time and cost savings. These studies should be considered with caution, however, because a strong interim analysis between the 2 and 3 phases of the study is mandatory, and in some instances patient accrual will have to be discontinued while the DSMC is making a decision.

Future Challenges

Designing clinical trials in HCC is particularly complex because of underlying nonmalignant liver disease. The guidelines developed by the AASLD expert panel and presented herein provide for the first time a working framework for designing trials with common rules that should facilitate comparability of results among researchers.

There remain several challenges in advancing the field of HCC clinical research. First, it should be established whether outcomes for a given tumor stage are equal in different geographic populations. Second, molecular classification might become clinically practical in the near future, and this might complicate trial design (6). The small number of centers and investigators currently involved in advanced clinical HCC research in comparison with other prevalent cancers is yet another challenge. Due to the complexity of the disease, the panel recommended that the design and conduct of phase 3 trials should be coordinated by centers with appropriate expertise in this specific cancer.

Minority Report

Dr A. X. Zhu has expressed concerns regarding the following. 1) Progression-free survival can be an acceptable endpoint in patients with well-preserved liver function because it will capture drug-induced safety signals. 2) CUPI (54) and CLIP (53) scoring systems are not discarded for trial design.

References

1. Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics 2002. *CA Cancer J Clin*. 2005;55(2):74–108.
2. Sangiovanni A, Del Ninno E, Fasani P, et al. Increased survival of cirrhotic patients with a hepatocellular carcinoma detected during surveillance. *Gastroenterology*. 2004;126(4):1005–1014.
3. El Serag HB, Mason AC. Rising incidence of hepatocellular carcinoma in the United States. *N Engl J Med*. 1999;340(10):745–750.
4. Llovet JM, Burroughs A, Bruix J. Hepatocellular carcinoma. *Lancet*. 2003;362(9399):1907–1917.
5. Llovet JM, Bruix J. Systematic review of randomized trials for unresectable hepatocellular carcinoma: chemoembolization improves survival. *Hepatology*. 2003;37(2):429–442.
6. López PM, Villanueva A, Llovet JM. Systematic review: evidence-based management of hepatocellular carcinoma—an updated analysis of randomized controlled trials. *Aliment Pharmacol Ther*. 2006;23(11):1535–1547.
7. Bruix J, Sherman M. Management of hepatocellular carcinoma. *Hepatology*. 2005;42(5):1208–1236.
8. Bruix J, Sherman M, Llovet JM, et al. Clinical management of hepatocellular carcinoma. Conclusions of the Barcelona-2000 EASL Conference. *J Hepatol*. 2001;35(3):421–430.
9. Llovet JM, Bru C, Bruix J. Prognosis of hepatocellular carcinoma: the BCLC staging classification. *Semin Liver Dis*. 1999;19(3):329–338.
10. Marrero JA, Fontana RJ, Barrat A, et al. Prognosis of hepatocellular carcinoma: comparison of 7 staging systems in an American cohort. *Hepatology*. 2005;41(4):707–716.
11. Cillo U, Vitale A, Grigoletto F, et al. Prospective validation of the Barcelona Clinic Liver Cancer staging system. *J Hepatol*. 2006;44(4):723–731.
12. Vauthey J, Lauwers G, Esnaola N, et al. Simplified staging for hepatocellular carcinoma. *J Clin Oncol*. 2002;20(6):1527–1536.
13. Llovet JM, Schwartz M, Mazzaferro V. Resection and liver transplantation for hepatocellular carcinoma. *Semin Liver Dis*. 2005;25(2):181–200.
14. Villa E, Moles A, Ferretti I, et al. Natural history of inoperable hepatocellular carcinoma: estrogen receptors' status in the tumor is the strongest prognostic factor for survival. *Hepatology*. 2000;32(2):233–238.
15. Mazzaferro V, Regalia E, Doci R, et al. Liver transplantation for the treatment of small hepatocellular carcinomas in patients with cirrhosis. *N Engl J Med*. 1996;334(11):693–699.
16. Lo CM, Ngan H, Tso WK, et al. Randomized controlled trial of transarterial lipiodol chemoembolization for unresectable hepatocellular carcinoma. *Hepatology*. 2002;35(5):1164–1171.
17. Llovet JM, Real MI, Montana X, et al. Arterial embolisation or chemoembolisation versus symptomatic treatment in patients with unresectable hepatocellular carcinoma: a randomised controlled trial. *Lancet*. 2002;359(9319):1734–1739.
18. Llovet J, Ricci S, Mazzaferro V, et al.; SHARP Investigators. Sorafenib improves survival in advanced Hepatocellular Carcinoma (HCC): results of a phase III randomized placebo-controlled trial. *J Clin Oncol*. 2007;LBA1.
19. Chien JY, Friedrich S, Heathman MA, de Alwis DP, Sinha V. Pharmacokinetics, pharmacodynamics, and the stages of drug development: role of modeling and simulation. *AAPS J*. 2005;7(3):E544–E559.

20. National Cancer Institute Web site. US National Institute of Health. http://www.cancer.gov. [Last accessed: April 17, 2008.]

21. Whitehead J. Stopping clinical trials by design. *Nat Rev Drug Discov.* 2004;3(11):973–977.

22. Pocock SJ. *Clinical Trials: A Practical Approach.* Chichester, UK: Wiley; 1983.

23. Jadad A, Moore A, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary?. *Control Clin Trials.* 1996; 17(1):1–12.

24. Moher D, Schulz KF, Altman DG: CONSORT Group. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet.* 2001;357(9263):1191–1194.

25. Wright JC, Weinstein MC. Gains in life expectancy from medical interventions-standardizing data on outcomes. *N Engl J Med.* 1998;339(6): 380–386.

26. Laupacis A, Feeny D, Detsky AS, et al. How attractive does a new technology have to be to warrant adoption and utilization? Tentative guidelines for using clinical and economic evaluations. *CMAJ.* 1992;146(4): 473–481.

27. Spruance SL, Reid JE, Grace M, Samore M. Hazard ratio in clinical trials. *Antimicrob Agents Chemother.* 2004;48(8):2787–2792.

28. Rothenberg ML, Carbone DP, Johnson DH. Improving the evaluation of new cancer treatments: challenges and opportunities. *Nat Rev Cancer.* 2003;3(4):303–309.

29. Lee JJ, Feng L. Randomized phase II designs in cancer clinical trials: current status and future directions. *J Clin Oncol.* 2005;23(19):4450–4457.

30. 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.

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

32. Kim WR, Therneau TM, Benson JT, et al. Deaths on the liver transplant waiting list: an analysis of competing risks. *Hepatology.* 2006;43(2):345–351.

33. Imamura H, Matsuyama Y, Tanaka E, et al. Risk factors contributing to early and late phase intrahepatic recurrence of hepatocellular carcinoma after hepatectomy. *J Hepatol.* 2003;38(2):200–207.

34. Cheung ST, Chen X, Guan XY, et al. Identify metastasis-associated genes in hepatocellular carcinoma through clonality delineation for multinodular tumor. *Cancer Res.* 2002;62(16):4711–4721.

35. Chen Y, Yeh SH, Chen JT, et al. Chromosomal changes and clonality relationship between primary and recurrent hepatocellular carcinoma. *Gastroenterology.* 2000;119(2):431–440.

36. 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.

37. Blazeby JM, Currie E, Zee BC, Chie WC, Poon RT, Garden OJEORTC 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.

38. Gennari A, Amadori D, De Lena M, et al. Lack of benefit of maintenance paclitaxel in first-line chemotherapy in metastatic breast cancer. *J Clin Oncol.* 2006;24(24):3912–3918.

39. Miller AB, Hoogstraten B, Staquet M, Winkler A. Reporting results of cancer treatment. *Cancer.* 1981;47(1):207–214.

40. Therasse P, Arbuck SG, Eisenhauer EA, et al. New guidelines to evaluate the response to treatment in solid tumors. *J Natl Cancer Inst.* 2000; 92(3):205–216.

41. Shepherd FA, Rodrigues Pereira J, et al.; National Cancer Institute of Canada Clinical Trials Group. Erlotinib in previously treated non-small-cell lung cancer. *N Engl J Med.* 2005;353(2):123–132.

42. Hudes G, Carducci M, Tomczak P, et al.; Global ARCC Trial. Temsirolimus, interferon alfa, or both for advanced renal-cell carcinoma. *N Engl J Med.* 2007;356(22):2271–2281.

43. Hurwitz H, Fehrenbacher L, Novotny W, et al. Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer. *N Engl J Med.* 2004;350(23):2335–2342.

44. Millar AW, Lynch KP. Rethinking clinical trials for cytostatic drugs. *Nat Rev Cancer.* 2003;3(7):540–545.

45. Kelloff GJ, Krohn KA, Larson SM, et al. The progress and promise of molecular imaging probes in oncologic drug development. *Clin Cancer Res.* 2005;11(22):7967–7985.

46. Hylton N. Dynamic contrast-enhanced magnetic resonance imaging as an imaging biomarker. *J Clin Oncol.* 2006;24(20):3293–3298.

47. Gharib AM, Thomasson D, Li KC. Molecular imaging of hepatocellular carcinoma. *Gastroenterology.* 2004;127(5 suppl 1):S153–S158.

48. Rebmann V, Schütt P, Brandhorst D, et al. Soluble MICA as an independent prognostic factor for the overall survival and progression-free survival of multiple myeloma patients. *Clin Immunol.* 2007;123(1):114–120.

49. Colli A, Fraquelli M, Casazza G, et al. Accuracy of ultrasonography, spiral CT, magnetic resonance, and alpha-fetoprotein in diagnosing hepatocellular carcinoma: a systematic review. *Am J Gastroenterol.* 2006;101(3):513–523.

50. Forner A, Vilana R, Ayuso C, et al. Diagnosis of hepatic nodules < 20 mm in cirrhosis. Prospective validation of the AASLD guidelines for HCC. *Hepatology.* 2008;47(1):97–104.

51. Llovet JM, Chen Y, Wurmbach W, et al. A molecular signature to discriminate dysplastic nodules and early hepatocellular carcinoma in HCV-cirrhosis. *Gastroenterology.* 2006;131(6):1758–1767.

52. 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(1):217–231.

53. 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.

54. 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.

55. Sala M, Llovet JM, Vilana R, et al. Initial response to percutaneous ablation predicts survival in patients with hepatocellular carcinoma. *Hepatology.* 2004;40(6):1352–1360.

56. Livraghi T, Giorgio A, Marin G, et al. Hepatocellular carcinoma and cirrhosis in 746 patients: long-term results of percutaneous ethanol injection. *Radiology.* 1995;197(1):101–108.

57. Kamath PS, Wiesner RH, Malinchoc M, et al. A model to predict survival in patients with end-stage liver disease. *Hepatology.* 2001;33(2):464–470.

58. Simon R, Thall PF, Ellenberg SS. New designs for the selection of treatments to be tested in randomized clinical trials. *Stat Med.* 1994; 13(5-7):417–429.

59. Slamon DJ, Leyland-Jones B, Shak S, et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med.* 2001;344(11):783–792.

60. Kobayashi S, Boggon TJ, Dayaram T, et al. EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. *N Engl J Med.* 2005; 352(8):786–792.

61. Lee JS, Heo J, Libbrecht L, et al. A novel prognostic subtype of human hepatocellular carcinoma derived from hepatic progenitor cells. *Nat Med.* 2006;12(4):410–416.

Funding

Notes