Hepatocellular Carcinoma: Reasons for Phase III Failure and Novel Perspectives on Trial Design

Josep M. Llovet1,2,3 and Virginia Hernandez-Gea1

Abstract

Hepatocellular carcinoma (HCC) is a major health problem. Most patients with HCC experience a recurrence after resection/ablation or are diagnosed at advanced stages. Sorafenib remains the only approved systemic drug for these patients. Molecular therapies targeting signaling cascades involved in hepatocarcinogenesis have been explored in phase III clinical trials. However, none of the drugs tested have shown positive results in the first-line (brivanib, sunitinib, erlotinib, and linifanib) or second-line (brivanib, everolimus) setting after sorafenib progression. Reasons for failure are heterogeneous and include lack of understanding of critical drivers of tumor progression/dissemination, liver toxicity, flaws in trial design, or marginal antitumoral potency. These trials are also challenging time to progression as a surrogate endpoint of survival. Trials ongoing testing drugs head-to-head versus sorafenib in "all comers" might have difficulties in achieving superior results in the first line. Novel trials are also designed testing drugs in biomarker-based subpopulations of patients with HCC. Most common mutations, however, are undruggable, such as p53 and CTNNB1. Two types of studies are proposed: (i) phase II pivotal proof-of-concept studies testing drugs blocking potential oncogenic addiction loops, such as the one testing MEK inhibitors in RAS+ patients or amplification of FGF19 as a target; and (ii) phase II to III studies using biomarker-based trial enrichment for defining HCC subpopulations, such as the case of enriching for MET-positive tumors. These strategies have been deemed successful in breast, melanoma, and lung cancers, and are expected to change the landscape of trial design of HCC. Clin Cancer Res; 20(8); 2072–9. ©2014 AACR.
Reasons for Failure of Phase III Studies

Hepatocellular carcinoma trials are characterized by patient heterogeneity and are complex to interpret and dissect (16). Our purpose is to provide some information on the reasons underlying the fact that phase III pivotal randomized controlled trials (RCT) in HCC reported during the past 2 years did not meet the primary endpoint, but full data have not yet been reported. We are herein focusing on exploring the causes of trial failure, and the concept of drivers, oncogenic addiction, and biomarker-driven trials will be discussed in more detail in a subsequent section of this article.

Table 1. Randomized phase III pivotal clinical trials completed in HCC in the first- and second-line settings (2007–2013)

<table>
<thead>
<tr>
<th>Comparison (reference)</th>
<th>Design</th>
<th>First-line advanced HCC</th>
<th>Second-line advanced HCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sorafenib vs. placebo (9) (SHARP trial)</td>
<td>Superiority</td>
<td>10.7 mo vs. 7.9 mo; HR = 0.69 (95% CI, 0.55–0.87); P = 0.00058</td>
<td>5.5 mo vs. 2.8 mo; HR = 0.58 (95% CI, 0.45–0.74); P = 0.001</td>
</tr>
<tr>
<td>Sorafenib vs. placebo (10) (Asia Pacific Trial)</td>
<td>Superiority</td>
<td>6.5 mo vs. 4.2 mo; HR = 0.68 (95% CI, 0.50–0.93); P = 0.014</td>
<td>2.8 mo vs. 1.4 mo; HR = 0.57 (95% CI, 0.42–0.79); P = 0.0005</td>
</tr>
<tr>
<td>Sunitinib vs. sorafenib (17) (SUN trial)</td>
<td>Superiority</td>
<td>7.2 mo vs. 10.2 mo; HR = 1.30 (95% CI, 1.13–1.5); P = 0.001</td>
<td>3.8 mo vs. 4.1 mo; HR = 1.13 (95% CI, 0.98–1.31); P = 0.16</td>
</tr>
<tr>
<td>Brivanib vs. sorafenib (18) (BRISK-FL trial)</td>
<td>Noninferiority</td>
<td>9.5 mo vs. 9.9 mo; HR = 1.05 (95% CI, 0.94–1.23); P = 0.31</td>
<td>4.1 mo vs. 4.2 mo; HR = 1.01 (95% CI, 0.88–1.16); P = 0.8</td>
</tr>
<tr>
<td>Linifanib vs. sorafenib (19) (LIGHT Trial)</td>
<td>Noninferiority</td>
<td>9.1 mo vs. 9.8 mo; HR = 1.04 (95% CI, 0.89–1.22); P = ns</td>
<td>5.4 mo vs. 4 mo; HR = 0.76 (95% CI, 0.64–0.89); P &lt; 0.001</td>
</tr>
<tr>
<td>Sorafenib ± erlotinib vs. sorafenib (20) (SEARCH trial)</td>
<td>Superiority</td>
<td>9.5 mo vs. 8.5 mo; HR = 0.92 (95% CI, 0.78–1.1); P = 0.2</td>
<td>3.2 mo vs. 4.0 mo; HR = 0.113 (95% CI, 0.94–1.36); P = 0.91</td>
</tr>
<tr>
<td>Brivanib vs. placebo (21) (BRISK-PS Trial)</td>
<td>Superiority</td>
<td>9.4 mo vs. 8.2 mo; HR = 0.89 (95% CI, 0.69–1.15); P = 0.33</td>
<td>4.2 mo vs. 2.7 mo; HR = 0.56 (95% CI, 0.42–0.78); P = 0.001</td>
</tr>
</tbody>
</table>

Abbreviation: TTP, time to progression.

aOverall survival (OS) is the primary endpoint in all studies.

bResults for everolimus vs. placebo have been reported as negative in a recent press release.
Lack of efficacy. The EGFR tyrosine kinase inhibitor erlotinib moved to phase III trial with modest disease control benefit in 2 previous phase II trials (26, 27). These single-arm phase II studies reported median survival of 10 to 13 months associated to marginal response rates (26, 27). There were no data testing the combination of sorafenib + erlotinib in phase II in HCC, but it was estimated that this combination would be able to improve OS in the SEARCH study (20). At the end of the study 720 patients were randomized being the median OS for the combination arm of 9.5 months compared with 8.5 months for the sorafenib arm (HR = 0.9; 95% CI, 0.78–1.1; P = 0.2). Considering that trial design, stratification, and execution of the trial was similar to SHARP, the results highlighted the fact that erlotinib is not adding outcome benefit in patients on sorafenib. It seems also that adding erlotinib led to cumulative adverse events decreasing sorafenib treatment duration from 4 to 3 months.

Linifanib, a more potent multikinase inhibitor targeting VEGFR and PDGFR than sorafenib, was tested in the first-line setting after a small single-arm phase II study including 44 patients showed median survival of 10.4 months (28). The phase III study with around 1,000 patients did not meet the noninferiority endpoint (OS = 9.1 versus 9.8 months; HR = 1.04; 95% CI 0.89–1.22, P < 0.001; ref. 19). The study was halted at the third interim analysis because of futility, although the reported data also suggest that linifanib was more toxic than sorafenib in terms of serious adverse events (54% vs. 38%), grade 3 to 4 hypertension (20% vs. 10%) and grade 3 to 4 liver-related complications (encephalopathy, ascites, and hyperbilirubinemia: 20% vs. 10%).

A better understanding of trial design: the noninferiority scenario. According to guidelines, new molecules tested in the first-line setting need to be combined with the standard of care sorafenib to demonstrate superiority (8, 16). This has only been the case in one RCT, sorafenib + erlotinib (20), while the others were directly testing the candidate drug versus sorafenib. We can estimate that in order to improve sorafenib outcome in the first-line setting for superiority, the drug has to be potent enough to achieve a HR of 0.5 or less if it is theoretically compared with the natural history of the disease or placebo (9, 29). This represents a very challenging scenario that requires a very potent drug able to decrease the risk of death by 50%. Nonetheless, two pivotal RCTs proposed a two-step design, first testing noninferiority and then testing for superiority (18, 19). Despite splitting the α between two analyses—which can be assumed increasing the sample size—it seems that this approach can give chances to drugs that show efficacy similar to that of sorafenib. However, the window of opportunity for these drugs is certainly too narrow.

Brivanib is a multikinase inhibitor with activity against fibroblast growth factor (FGF) signaling that achieved a 10-month OS (95% CI, 6.8–15.2) with a good tolerance and safety profile in a first-line single-arm 55-patient trial (30). The phase III pivotal study was designed as a noninferiority trial under the assumption that brivanib would have equal effects to sorafenib or would not be unacceptably worse with manageable adverse events, and with near absence of liver toxicity or treatment-related deaths (9, 10). Because of the success of the SHARP trial testing sorafenib (9), its design has been adopted by almost all studies testing molecular therapies in the first-line setting. The main concepts were adopted by guidelines of trial design (8, 16) and are as follows: (i) select only patients with preserved liver function (Child-Pugh A). Only patients with Child-Pugh A should be included in trials to prevent liver failure and death as a result of the natural history of cirrhosis. Patients with Child-Pugh A without HCC are expected to have a 1-year survival rate of 95%, a percentage that is of utmost importance to prevent competing risks for death (the main endpoint). Underlying liver disease highly complicates HCC treatment, as it can be itself responsible for morbidity and mortality. (ii) Select patients with advanced-stage HCC according to the Barcelona Clinic Liver Cancer (BCLC-C) staging system (2, 4, 22), or with BCLC-B progressing after the standard of care for intermediate cases, chemotherapy. Other classifiers of patients—such as CLIP (Cancer of the Liver Italian Program) score (23)—might not be appropriate for treatment allocation, because the primary purpose of the score was to define prognosis but not to allocate for specific therapies. (iii) Define OS as primary endpoint, and time to progression (TTP) as secondary end point. Progression-free survival is a vulnerable endpoint in HCC research (16).

Sunitinib, a drug similar to sorafenib in terms of the inhibitory kinase profile—except for B-Raf inhibition, which is important in HCC signaling—and with better antitumor activity [lower 50% inhibitory concentration (IC50) for almost all of the targets] was approved by the FDA in 2006 for the treatment of renal carcinoma and gastrointestinal stromal tumors. Nonetheless, sunitinib did not meet the survival superiority endpoint versus sorafenib in a phase III study (17) probably as a result of two factors. First, signals of efficacy and toxicity were already present in phase II studies (24, 25). Second, head-to-head comparisons with the standard of care (sorafenib) should only be conducted if the signal of efficacy in phase II is robust. Two phase II single-arm trials testing sunitinib showed liver toxicity (encephalopathy, hepato-renal syndrome) leading to death in the range of 6% to 10% (24, 25). Both studies including less than 40 patients showed median survival in the range of 8 to 10 months, associated with 10% (50 mg/day; ref. 24) and 6% (37.5 mg/day; ref. 25) treatment-related death, respectively. Based on these data, 37.5-mg dose was used in a subsequent phase III study randomizing 1,074 patients, which was halted for toxicity/futility after a data safety watch at 457 events (17). Results at that point showed median survival of 7.9 months versus 10.2 months (sunitinib vs. sorafenib; HR = 1.30). Treatment-related deaths accounted for 3.2% of cases (17/526 patients) in the sunitinib arm versus 0.3% (2/542) in the sorafenib arm. Subgroup analysis showed higher benefit of sorafenib for hepatitis C virus–related patients with HCC (median survival of 15 months).
than sorafenib (assuming a decrease in OS no more than 3 weeks; ref. 18). Figure 1 shows the results of the phase III trial including 1,155 patients where the boundary of the confidence interval of the HR for survival was set at 1.08 following recommendations for noninferiority trial design. In summary, it is estimated that RCTs comparing drugs head-to-head to sorafenib in the first-line setting might have three potential results: (i) the drug is superior to sorafenib, if the HR (95% CI) boundaries do not cross the unity; (ii) the drug is noninferior compared with sorafenib, if the HR (95% CI) boundaries fall between 1 and 1.08; and (iii) the drug is inferior to sorafenib if the HR (95% CI) boundaries cross the 1.08 upper limit for noninferiority. The figure shows the HR and 95% CI of two RCTs, showing that according to this design the drugs are inferior to sorafenib (18, 19).

Lessons learned

After analyzing the results of these trials, we have learned some lessons. First, specific and properly powered phase II studies exploring liver-related toxicity are required in cirrhotic patients with HCC prior to testing drugs in phase III. In fact, liver-induced toxicity might hamper the positive results obtained by potent antitumoral drugs, as can be the case with sunitinib, in which drug-related liver adverse events might have counterbalanced the antitumoral benefits (17). In addition, it has been proposed that these phase II trials incorporate a control arm that would reduce selection bias and allow impartial comparison in secondary analysis (16, 32, 33). This strategy might have diminished the risk of a negative phase III trial. It has to be acknowledged, however, that the positive signal leading to the sorafenib phase III trial was obtained in a large single-arm phase II study including more than 100 patients (34). Large phase II studies might prevent biased results because of the regression-to-the-mean correction effect by which the higher the sample, the lower the risk of random error in estimating a survival outcome. Thus, this approach remains as a valid alternative to randomized phase II studies. Second, trial stratification in the second-line setting would fit better to test drugs in liver oncology as it may include a control arm that would reduce selection bias and allow impartial comparison in secondary analysis.

Figure 1. Understanding noninferiority study design and HR boundaries in HCC research. It is estimated that RCTs comparing drugs head-to-head to sorafenib in the first-line setting might have three potential results: (i) the drug is superior to sorafenib, if the HR (95% CI) boundaries do not cross the unity; (ii) the drug is noninferior compared with sorafenib, if the HR (95% CI) boundaries fall between 1 and 1.08; and (iii) the drug is inferior to sorafenib if the HR (95% CI) boundaries cross the 1.08 upper limit for noninferiority. The figure shows the HR and 95% CI of two RCTs, showing that according to this design the drugs are inferior to sorafenib (18, 19).
Novel Perspectives on Trial Design

Most phase II and III studies currently ongoing in HCC are targeting "all comers." Such is the case of levatinib, Y-90, doxorubicin, or oncolytic virus, which are challenging sorafenib in the first-line setting, and ramucirumab, regorafenib, or cabozantinib versus placebo in second line. Although this strategy remains valid and probably will be the most common approach in the field, it is expected that an increasing number of studies targeting specific molecular subclasses of HCC will emerge. To understand the rationale for trial design based on molecular markers, we first review the molecular classification and drivers as potential targets of therapies in HCC reported so far and then explore the two proposed trial designs: proof-of-concept trials and biomarker-based enrichment trials.

Overview of molecular classification and drivers as targets for therapies in HCC

Molecular classification of cancer should aid in understanding the biologic subclasses and drivers of the disease, optimize benefits from molecular therapies, and enrich trial populations. In HCC, no molecular subclass has been reported as responding to a specific targeted therapy so far (8). From the biologic standpoint, different tumor classes have been characterized, including a Wnt subclass (enriched with CTNNB1 mutations), a proliferation class (with two subcategories: S1-TGF-β and S2-EpCAM positive) and an inflammation class (37–40). The proliferation subclass accounts for 50% of cases and is enriched with tumors derived from progenitor cells (e.g., "EpCAM"); ref. 38), and these cases tend to have worse prognosis. In addition, RAS, IGF, MET, and mTOR signaling are pathways clearly enriched in this subclass (37).

High-resolution analysis of molecular alterations in human malignancies has allowed the identification of new disease drivers to allow the modification of treatments in some solid malignancies such as lung, breast, or melanoma. Recent studies have provided a broad picture of the mutational profile in HCC and identified an average of 30 to 40 mutations per tumor, among which six to eight are considered drivers (41, 42). The main mutations described are in the promoter region of TERT, p53, CTNNB1, ARID1A, and Axin 1 (see Table 2; ref. 42). Deep-sequencing studies confirmed TP53 and CTNNB1 as frequently mutated in HCC (41, 42). Of note, mutations in these genes are mutually exclusive—an indication that they could act as drivers of tumor progression. In addition, these studies revealed novel mutations associated with HCC in different members of the chromatin remodeling pathway (ARID1A and ARID2), in genes involved in ubiquitination (KEAP1), RAS/MAPK signaling (RPS6KA3), and oxidative stress (NFE2L2), and JAK1 in 9% of hepatitis B virus (HBV)-related HCC. Functional validation of these novel mutations in experimental models suggests that JAK1 inhibition represents an attractive new therapeutic target. Furthermore, it is clear that genes commonly mutated in other solid tumors such as EGFR, Braf, Pik3Ca, or KRAS are rarely mutated in HCC (<5% of cases; Table 2; ref. 42). High-level amplifications at 5% to 10% prevalence containing oncogenes have been described in 11q13 (Cyclin D1 and FGF19) and 6p21 (VEGFA; ref. 37), whereas other more common amplifications reported contain Myc and Met genes. Some of these confirmed drivers represent druggable targets for therapies.

Proof-of-concept trials and trial enrichment

We can distinguish two types of biomarkers defining different trial designs: oncogenic addition loops, defining proof-of-concept trials (the case of refametinib in RAS-positive mutants in phase II; ref. 43), and the activation of signaling cascades leading to trial enrichment based on biomarkers (the case of tivantinib in patients with MET-positive HCC in phase II and III; ref. 44).

Pivotal proof-of-concept phase II trials. Several molecular therapies have been approved in oncology during the past decade targeting oncogenic addiction loops. Oncogene addiction can be defined as a heightened state of dependency of cancer cells on activated oncogenes or loss of...
tumor suppressors. Therefore, not all genetic alterations found in cancer genomes are able to establish this dependence. Numerous human cancers highlight the therapeutic perspective of oncogene addiction. For instance, the kinase activity of the BCR-ABL fusion protein is implicated in the pathogenesis of chronic myeloid leukemia, and once this loop is effectively discontinued with imatinib, tumor burden decreases and there is a positive effect in patient survival (45). Similarly, EGFR mutations in lung cancer predict responses to erlotinib/gefitinib (46), and patients with amplification in HER2/neu in breast cancer respond to trastuzumab, a monoclonal antibody directed to this receptor (47). More recent examples are melanoma tumors harboring B-RAF mutations responding to vemurafenib (48) and ALK fusions responding to crizotinib in NSCLC. In these latter cases, proof-of-concept pivotal single-arm trials led to accelerated FDA approval.

Despite being an attractive strategy, there are some issues that need to be addressed when implementing this concept in HCC: (i) discovery of bona fide oncogenic addiction loops; (ii) selection of adequate ad hoc blockade drugs. For instance, WNT activation occurs in 50% of cases (49), but WNT inhibitors have not yet entered advanced clinical developmental phases; and (iii) HCC has a significant genomic heterogeneity, which facilitates the coexistence of competing loops within the same tumor. Despite these drawbacks, it is time to develop personalized medicine approaches in HCC. A few "druggable" drivers of tumor progression and metastasis have been identified so far, such as RAS, FGF19, IGF2, NOTCH1, MET and mTOR signaling (42, 50). An example of a proof-of-principle early trial that is ongoing is the one testing the MEK inhibitor refametinib in specific RAS-mutated HCC cases (~5% of patients with HCC). In this trial, 300 patients with HCC will be screened in order to treat around 15 RAS-positive patients (43).

**Trial enrichment for activation of signaling pathways.** Validated biomarkers predicting response to sorafenib have not been identified. High c-KIT or low HGF were identified as predictors of survival in patients treated with sorafenib, thus raising the possibility that these markers could identify better drug responders (51). However, none of them ultimately showed predictive value (P of interaction: 0.081 and 0.073, respectively; ref. 51). Further efforts should be focused in enriching trial populations with biomarkers defining molecular subclasses. This can be the case with some pathways that are clearly activated in HCC, such as mTOR, IGF, MET, and WNT among others. One can speculate that the reasons why trial enrichment has not been implemented in HCC research until very recently might be related to (i) difficulties in obtaining tumor tissue in patients already diagnosed by noninvasive criteria; (ii) lack of effective biomarkers to be applied for the drugs tested so far in phase III; and (iii) limited understanding of the pathogenesis of the disease and primary drivers in selected populations.

Recently, a phase II trial comparing tivantinib (a selective oral inhibitor of MET) versus placebo was conducted in the second-line setting. No effect was identified in "all comers," but a significant survival effect was achieved in the post hoc analysis of MET-positive patients (OS = 7.2 vs 3.8; HR = 0.38;
95% CI, 0.18–0.81; \( P = 0.01; \) ref. 44). Confirmation of results is ongoing in phase III. From the biologic standpoint, adverse events related to tivantinib, such as neutropenia, suggest that the drug has off-target effects. Another c-Met inhibitor, cabo-zatinib, is being tested in phase III. This drug is a multitarget inhibitor also targeting VEGF, and the trial is conducted in the second line in all patients progressing to sorafenib.

We are facing a new era for testing drugs in HCC. On one hand, nonspecific drugs that target all patients will still be explored, and there is room for such types of drugs particularly in the second-line (this is the case of ramucirumab and regorafenib) or third-line setting, when there is such unmet need. On the other hand, pivotal proof-of-concept trials or trials with biomarker-based enrichment will emerge for specific pockets of HCC populations, which can completely change the treatment paradigm not only of advanced cases, but also of early or intermediate stages. In parallel, some etiology-specific activated pathways or oncogenes might be discovered or confirmed in the near future and should be considered for further modeling of trial design. Thus, the dawn of a new era of HCC trials is coming.

Disclosure of Potential Conflicts of Interest

J.M. Llovet is a consultant/advisory board member for Bayer Pharmaceuticals, Boehringer/Ingelheim, Bristol-Myers Squibb, Blueprint, Lilly-Imclone, Novartis, and Jennerex. No potential conflicts of interest were disclosed by the other author.

Received July 10, 2013; revised January 21, 2014; accepted February 5, 2014; published OnlineFirst March 3, 2014.

References


43. Refametinib in Ras mutant HCC. In Clinical trials.gov. (Identifier: NCT01915589).
Hepatocellular Carcinoma: Reasons for Phase III Failure and Novel Perspectives on Trial Design

Josep M. Llovet and Virginia Hernandez-Gea


Updated version
Access the most recent version of this article at:
doi:10.1158/1078-0432.CCR-13-0547

Cited articles
This article cites 48 articles, 16 of which you can access for free at:
http://clincancerres.aacrjournals.org/content/20/8/2072.full.html#ref-list-1

Citing articles
This article has been cited by 11 HighWire-hosted articles. Access the articles at:
/content/20/8/2072.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.