Linifanib versus Sorafenib in Patients with Advanced Hepatocellular Carcinoma: Results of a Randomized Phase III Trial

Cainap, et al

S1. Statistical Methods

Unless otherwise noted, for all statistical analysis, statistical significance was determined by a 2-sided p value ≤ 0.05 (when rounded to 3 decimal places).

S1.1 Sample Size Determination

Assuming the true hazard ratio (HR) in favor of the linifanib group is 0.80, a total of 667 deaths would be needed for the study to have 80% power at a 1-sided α level of 0.025 to detect a statistically significant treatment effect for the linifanib group using the log-rank test for overall survival (OS). Two interim analyses, one for futility only and one for both efficacy and futility, were performed and reviewed by an Independent Data Monitoring Committee (IDMC) when approximately 200 deaths (30% of the required number of events) and 333 deaths (50% of the required number of events) were observed, respectively. The Lan-DeMets alpha spending function with an O'Brien-Fleming boundary was to be used to ensure that the 1-sided false positive rate would be 0.025 or less for OS.

Further assuming that the time to death in both treatment groups would follow an exponential survival distribution, a total of approximately 900 patients were to be enrolled into the study during a 20-month period, with a maximal follow-up period of 35 months.

The power of the study (with 667 events) corresponding to a range of possible true HRs is tabulated below.

Power (for Superiority Test)	True HR
80%	0.80
68%	0.82
54%	0.84
46%	0.85

Two key sorafenib Phase III trials were conducted to assess the efficacy of sorafenib versus placebo in patients with advanced HCC: the SHARP trial[1] and the Asia-Pacific trial[2]. In the SHARP trial[1], which included 602 HCC patients from Europe, Australasia, North America, and Central and South America, the estimated HR (95% CI) for sorafenib versus placebo was 0.69 (0.55, 0.87). In the Asia-Pacific trial[2], which included 226 HCC patients from China, South Korea, and Taiwan, the estimated HR (95% CI) for sorafenib versus placebo was 0.68 (0.50, 0.93). Using a meta-analysis method proposed by Whitehead et al[3] and Parmar MK et al[4], the pooled HR and the 95% CI were 0.6865 (0.5709, 0.8255). Using the 95% CI lower limit method on log HR as described in Rothmann et al[5], the non-inferiority margin (or cutoff) corresponding to a 75% retention of the sorafenib effect was calculated as 1.0491.

Using a non-inferiority test on the primary efficacy endpoint of OS with a non-inferiority margin of 1.0491, the power of the study to declare non-inferiority is tabulated below for the same range of possible HRs (one interim analysis for futility only and one interim analysis for both efficacy and futility are assumed when approximately 200 and 333 deaths occur, respectively).

Power (for Non-Inferiority Test)	True HR
93%	0.80
88%	0.82
80%	0.84
74%	0.85

S1.2 Interim Analysis

A total of two interim analyses were performed.

The first interim analysis was performed for futility only, using OS as the endpoint. It was performed when approximately 200 deaths (30% of the required events) were observed. The Lan-DeMets alpha spending function was used to derive the futility stopping boundary.

The second interim analysis was performed for both efficacy and futility, using OS as the endpoint. It was performed when approximately 333 deaths (50% of the required events) were observed. The Lan-DeMets alpha spending function with an O'Brien-Fleming boundary was used to ensure that the 1-sided false positive rate was 0.025 or less for OS.

Interim statistical analyses and summaries for presentation to the IDMC were prepared by a clinical research organization. AbbVie personnel did not have access to the interim analyses prepared for the IDMC.

In making any recommendation regarding discontinuation of the study (either for efficacy or futility) at the interim analyses, the IDMC was guided by a formal stopping rule based on the primary efficacy endpoint of OS. The O'Brien-Fleming boundaries at the interim analyses and the final analysis are tabulated below. The derivation of the futility boundary is based on a non-inferiority design, with an assumed true HR of 0.80 and a non-inferiority margin of 1.0491.

Analysis	Number of Observed Events	Efficacy Stopping Boundary 1-sided p value	Futility Stopping Boundary 1-sided p value
First Interim Analysis	200	N/A	≥0.8981
Second Interim Analysis	333	≤0.0015	≥0.4646
Final Analysis	667	≤0.0245	

N/A=not applicable.

At the planned interim analysis, if the study demonstrated a statistically significant and clinically meaningful improvement in OS as compared with sorafenib, the study was to be stopped for this overwhelming efficacy. The data cutoff date for the final primary efficacy analyses would then be the data cutoff date of the interim analysis.

S1.3 Statistical Methods for Efficacy Analyses

The primary efficacy analysis was a comparison of OS distributions between the linifanib and sorafenib treatment groups.

Time to death (ie, OS) for a given patient was defined as the number of days from the date that the patient was randomized to the date of the patient's death. All events of death on or before the "cutoff" date (the date of the 667th death) were included, regardless of whether the patient died while still taking the study drug or after discontinuing the study drug. The distribution of OS was estimated for each treatment group using Kaplan-Meier methodology. Estimated median survival time and 95% CI for the estimated median survival time were presented for each treatment group. Both non-inferiority and superiority hypotheses were tested for the primary efficacy endpoint of OS using the Cox proportional hazard model with treatment as a factor, stratified by region (non-Asia, Japan, or rest of Asia), Eastern Cooperative Oncology Group (ECOG) performance status (0 vs. 1), vascular invasion or extrahepatic spread (yes vs.

no), and hepatitis B virus infection (yes vs. no). Non-inferiority for OS was tested first with a margin value of 1.0491. If non-inferiority was declared for OS, then superiority was to be tested for OS. The HR and the corresponding CI were estimated using the stratified Cox proportional hazard model.

If the study was not stopped for superior efficacy at the interim analysis, non-inferiority was to be declared at the final analysis if the upper limit of the 2-sided 95.1% CI for HR was ≤1.0491, which would mean that linifanib preserved at least 75% of the sorafenib treatment effect as observed in the sorafenib SHARP[1] and Asia-Pacific[2] trials. Furthermore, superiority was to be declared at the final analysis if the upper limit of the 2-sided 95.1% CI for HR was ≤1.0.

For a given patient, time to progression (TTP) was defined as the number of days from the day the patient was randomized to the first day the patient experienced an event of disease progression, per RECIST version (v) 1.1. All events of disease progression occurring on or before the "cutoff" date were included, regardless of whether the event occurred while the patient was still taking the study drug or had previously discontinued the study drug. If the patient did not have an event of disease progression on or before the "cutoff" date, the patient's data were censored at the date of the patient's last available radiographic tumor assessment on or before the "cutoff" date, with the following exception: if there was any radiographic tumor assessment after the "cutoff" date demonstrating evidence of the patient being progression-free per RECIST v1.1, the censoring date was set to the "cutoff" date. If a patient did not have any post-baseline radiographic tumor assessment, the patient was censored on the date of randomization.

TTP was analyzed using the same statistical method as for OS.

The objective response rate based on radiographic tumor assessment per RECIST v1.1, defined as the proportion of patients with a confirmed complete or partial response, was compared between the two treatment groups using the Cochran-Mantel-Haenszel (CMH) test, stratified by region (non-Asia, Japan, or rest of Asia), ECOG performance status (0 vs. 1), vascular invasion or extrahepatic spread (yes vs. no), and hepatitis B virus infection (yes vs. no). In addition, the objective response rate and the corresponding 95% CI were obtained for each treatment group. All patients who were randomized, regardless of whether they have any post-baseline radiographic tumor assessment, were included in the analysis.

For a given patient, progression-free survival (PFS) was defined as the number of days from the day the patient was randomized to the first day the patient experienced an event of disease progression per RECIST v1.1 or to the date of death (all causes of mortality) if disease progression was not reached. All events of disease progression occurring on or before the "cutoff" date were included, regardless of whether the event occurred while the patient was still taking the study drug or had previously discontinued the study drug. All events of death occurring on or before the "cutoff" date were included for patients who had not experienced disease progression, provided the death occurred within 42 days of the last radiographic tumor assessment. If the patient did not have an event of disease progression and had not died on or before the "cutoff" date, the patient's data were censored at the date of the patient's last available radiographic tumor assessment on or before the "cutoff" date, with the following exception: if there was any radiographic tumor assessment after the "cutoff" date demonstrating evidence of the patient being progression-free per RECIST v1.1, the censoring date was set to the "cutoff" date. If the randomized patient did not have any post-baseline radiographic tumor assessment, the patient's data were censored on the date of randomization.

PFS was analyzed using the same methodology as for OS and TTP.

S1.4 Statistical Methods for Safety Analyses

Analyses of adverse events (AEs) included only "treatment-emergent" events. "Treatment-emergent AEs" were defined as any AEs that first occurred on or after the date of first dosing and with an onset date no more than 30 days after the last dose of study drug. Treatment-emergent AEs were summarized by system organ class and preferred term according to the MedDRA AE coding dictionary. The percentage of patients experiencing an AE at a NCI CTCAE toxicity grade and the relationship of the AE to the study drug were provided. Serious AEs, AEs leading to study drug discontinuation due to disease progression and not due to disease progression, AEs leading to study drug interruption, and AEs leading to study drug dose reduction were summarized and compared between the two treatment groups using Fisher's exact test.

S1.5 Variables Used for Stratification of Randomization

Patient randomization was stratified by region (non-Asia, Japan, or rest of Asia), ECOG performance status (0 vs. 1), vascular invasion or extrahepatic spread (yes vs. no), and hepatitis B virus infection (yes vs. no).

Randomization schedules were generated using stratified permuted blocks with mixed block size of 2 or 4.

References

1. Llovet JM, Ricci S, Mazzaferro V, et al: Sorafenib in advanced hepatocellular carcinoma. N Engl J Med 359:378-390, 2008
2. Cheng AL, Kang YK, Chen Z, et al: Efficacy and safety of sorafenib in patients in the Asia-Pacific region with advanced hepatocellular carcinoma: a phase III randomised, double-blind, placebo-controlled trial. Lancet Oncol 10:25-34, 2009
3. Whitehead A, Whitehead J: A general parametric approach to the meta-analysis of randomized clinical trials. Stat Med 10:1665-1677, 1991
4. Parmar MK, Torri V, Stewart L: Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. Stat Med 17:2815-2834, 1998
5. Rothmann M, Li N, Chen G, et al: Design and analysis of non-inferiority mortality trials in oncology. Stat Med 22:239-264, 2003