Which Response Criteria Best Help Predict Survival of Patients with Hepatocellular Carcinoma Following Chemoembolization?

A Validation Study of Old and New Models[1]

Ju Hyun Shim, MD
Han Chu Lee, MD, PhD
Seon-Ok Kim, MS
Yong Moon Shin, MD, PhD
Kang Mo Kim, MD, PhD
Young-Suk Lim, MD, PhD
Dong Jin Suh, MD, PhD

Purpose: To identify differences in radiologic assessment methods and determine optimal imaging criteria for response evaluation in hepatocellular carcinoma (HCC) patients treated with chemoembolization.

Materials and Methods: Institutional review board approval was obtained, and patient informed consent was waived. The present study included 332 patients with intermediate stage HCC and Child-Pugh A cirrhosis who underwent serial chemoembolization. All measurable target lesions of 1 cm or larger in diameter were uni- and bidimensionally measured both at baseline and during follow-up. Intermodel agreement among the guidelines of the World Health Organization (WHO), Response Evaluation Criteria in Solid Tumors (RECIST), the European Association for the Study of the Liver (EASL), and modified RECIST (mRECIST) were examined. The most reliable model was selected on the basis of the correlation with survival prediction.

Results: The κ values of comparisons among WHO, RECIST, and mRECIST guidelines were less than 0.20, whereas the κ value for the comparison of EASL and mRECIST guidelines was 0.94. In patients with a partial response (PR), stable disease (SD), or progressive disease (PD), compared with patients with a complete response (CR), hazard ratios (HRs) for survival were 2.99 (95% confidence interval [CI]: 2.14, 4.17), 3.49 (95% CI: 1.71, 7.10), and 15.63 (95% CI: 9.51, 25.69), respectively, for EASL criteria. In patients with a PR, SD, or PD, compared with patients with a CR, the HRs were 2.75 (95% CI: 1.96, 3.87), 6.32 (95% CI: 3.67, 10.90), and 16.06 (95% CI: 9.76, 26.43), respectively, for mRECIST guidelines ($P < .001$). The *C* index for the multivariate model was 0.76 (95% CI: 0.72, 0.79) for both EASL and mRECIST guidelines, thus exhibiting satisfactory capability to help predict survival. The Cox regression model revealed that both mRECIST and EASL guidelines were independent predictors of overall survival ($P < .001$ for both).

Conclusion: The enhancement models more accurately helped predict long-term survival in HCC patients treated with chemoembolization.

[1] From the Departments of Internal Medicine (J.H.S., H.C.L., K.M.K., Y.S.L., D.J.S.) and Radiology (Y.M.S.), Asan Liver Center, and Department of Biostatistics (S.O.K.), Asan Medical Center, University of Ulsan College of Medicine, 388-1 Poongnap-2dong, Songpa-gu, Seoul 138-736, Korea. Received February 8, 2011; revision requested April 6; final revision received June 24; accepted August 23; final version accepted September 13. **Address correspondence to** H.C.L. (e-mail: *hch@amc.seoul.kr*).

Chemoembolization is a standard component of local-regional care in nonsurgical patients with intermediate-stage hepatocellular carcinoma (HCC) defined as Barcelona Clinic Liver Cancer (BCLC) stage B. This condition is defined by multinodular asymptomatic tumors without vascular involvement or extrahepatic spread (1,2). A previous clinical trial has shown that objective tumor response following chemoembolization correlates well with survival (3).

Conventional response assessment guidelines for treatment of solid tumors (eg, the World Health Organization [WHO] criteria [4] or the Response Evaluation Criteria in Solid Tumors [RECIST] evaluation [5]) do not consider ischemic necrosis in tumors induced by transarterial intervention but rather focus on shrinkage of the entire tumor. To correct for underestimation of the real response rate with use of the original size criteria, clinical trials are required to explore standardized response criteria specific for local-regional therapeutics in patients with HCC. Such criteria were introduced by the HCC panel of the European Association for the Study of the Liver (EASL) (2). The newly proposed criteria, with which a bidimensional approach that is based on the original WHO criteria is used, assess only viable target tumors (ie, those defined as showing contrast material–enhancing areas in the arterial phase of a dynamic scan). Thereafter, earlier necrosis guidelines have been recently modified, employing a simple arithmetic method with which a single-plane assessment is used to measure the effectiveness of local-regional therapies in patients with HCC. The revised guidelines are known as the modified RECIST (mRECIST). Quantitative evidence is required to establish mRECIST assessment as a true surrogate for measurement of overall survival (OS), and to determine whether mRECIST is superior to traditional or current guidelines in estimation of chemoembolization responses in patients with HCC.

Given the need for clinical validation of any updated guidelines, we sought to identify differences in the various radiologic assessment methods, correlating response with survival, and to determine optimal criteria for evaluation of response outcome data in HCC patients treated with chemoembolization. To this end, we conducted response measurements according to WHO, RECIST, EASL, and mRECIST guidelines, by using a group of BCLC stage B patients of equivalent liver function and tumor status, who were accepted as appropriate candidates for chemoembolization.

Advances in Knowledge

- The modified Response Evaluation Criteria in Solid Tumors (mRECIST) enhancement guidelines showed good intercriterion agreement with the European Association for the Study of the Liver guidelines, in contrast with the conventional size criteria.
- The enhancement criteria more reliably helped predict long-term survival in hepatocellular carcinoma (HCC) patients treated with chemoembolization than other size-based imaging guidelines.

Materials and Methods

Study Design and Data Sources

Of the 1351 HCC patients initially treated with chemoembolization between 2000 and 2007 at Asan Medical Center (Seoul, Korea), we retrospectively selected 332 patients with both well-preserved liver function without ascites (Child-Pugh class A) and intermediate-stage multifocal HCC without any symptoms (BCLC stage B) prior to treatment (Fig 1). This eliminated potential confounders for survival, because in most patients with HCC, the cancer itself and the underlying liver disease each independently determine final patient outcome (6). No patient also had a history of other malignant disease, or of uncontrolled functional or metabolic disease, that could influence survival time. At initial diagnosis, all patients had at least one index lesion measuring 1 cm or larger in diameter (ie, the target lesion), wherein the typical features of HCC of arterial enhancement followed by washout during the portal venous phase could be observed on a dynamic scan, and in these patients, the lesions were confirmed as HCC, even without histologic evaluation that was based on American Association for the Study of Liver Diseases or EASL guidelines (1,2). Target lesions

Implications for Patient Care

- The enhancement criteria are expected to provide a valuable measure of treatment response in HCC clinical trials with use of chemoembolization and could be extrapolated to explore the response to other local-regional interventions.
- The mRECIST guidelines can be rapidly and effortlessly applied to clearly evaluate tumor responses to chemoembolization in clinical practice.

Published online before print
10.1148/radiol.11110282 **Content codes:** GI IR

Radiology 2012; 262:708–718

Abbreviations:
BCLC = Barcelona Clinic Liver Cancer
CI = confidence interval
CR = complete response
EASL = European Association for the Study of the Liver
HCC = hepatocellular carcinoma
HR = hazard ratio
IQR = interquartile range
mRECIST = modified RECIST
OS = overall survival
PD = progressive disease
PR = partial response
RECIST = Response Evaluation Criteria in Solid Tumors
SD = stable disease
TACE = transarterial chemoembolization
TTP = time to progression
WHO = World Health Organization

Author contributions:
Guarantors of integrity of entire study, J.H.S., H.C.L.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; literature research, J.H.S., H.C.L.; clinical studies, J.H.S., H.C.L., Y.M.S., K.M.K., Y.S.L., D.J.S.; experimental studies, J.H.S., H.C.L., S.O.K.; statistical analysis, J.H.S., H.C.L.; and manuscript editing, J.H.S., H.C.L., D.J.S.

Potential conflicts of interest are listed at the end of this article.

Figure 1

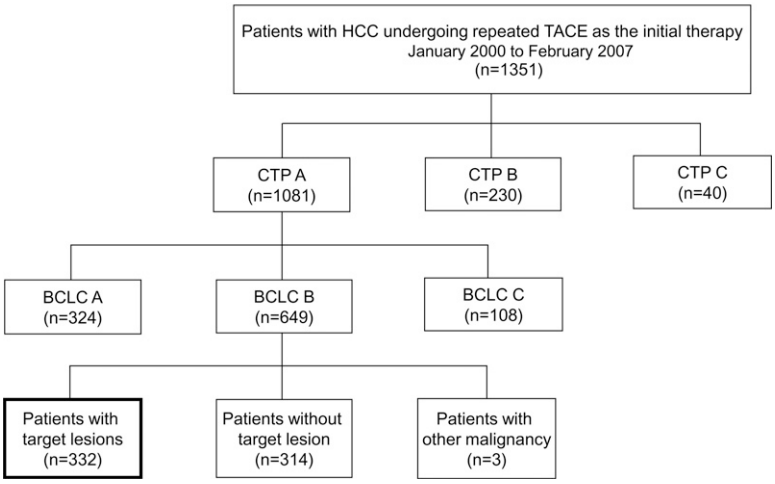

Figure 1: Flowchart of the patient selection process. *CTP A* = Child-Turcotte-Pugh class A, *CTP B* = Child-Turcotte-Pugh class B, *CTP C* = Child-Turcotte-Pugh class C, TACE = transarterial chemoembolization.

were characterized as distinctly nodular and not infiltrative, thus permitting accurate repetitive measurement (7). The remaining lesions other than target lesions for each method always included small lesions with a maximum diameter of less than 1 cm, and truly nonmeasurable lesions were considered as nontarget lesions.

This study was approved by the institutional review board, which has been officially recognized by the Forum for Ethical Review Committees in Asia and the Western Pacific, of our hospital.

Transarterial Intervention

The chemoembolization procedure used in our institution has been described elsewhere (8). Briefly, both superior mesenteric and common hepatic arteriography were performed to assess overall anatomy, tumor burden, and portal vein patency. Cisplatin (Cisplan; Dong-A Pharmaceutical, Seosan, Korea) was then infused into the lobar hepatic artery for 15 minutes without embolic particle administration. The infused dose of cisplatin was 2 mg per kilogram of body weight. After selective catheterization of the feeding artery with a microcatheter, an emulsion of 2–20 mL of iodized oil (Lipiodol Ultra-Fluide; Laboratoires Guerbet, Aulnay-sous-Bois, France) and cisplatin in a 1:1 ratio was infused into the feeding arteries. The feeder arteries were subsequently embolized by using 1-mm-diameter absorbable gelatin sponge particles (Gelfoam; Upjohn, Kalamazoo, Mich) until arterial flow stasis was achieved.

Repeated chemoembolization was indicated every 6–8 weeks if residual viable tumor tissue was evident on sequential triphasic computed tomography (CT) without the appearance of extrahepatic metastases, major portal vein invasion, or deterioration in liver function. Informed consent for chemoembolization was obtained from each patient prior to the commencement of any procedure.

Pre- and Postprocedural Work-up

Prior to the initial chemoembolization session, patients underwent laboratory tests, including a liver function panel, α-fetoprotein level, and hepatitis serologic tests, as well as dynamic liver CT and a metastatic work-up. Follow-up CT was performed 1 month after each chemoembolization session to assess response to treatment and to allow a timely decision on subsequent treatment. When indicated, 88 (26.5%) of 332 patients underwent dynamic magnetic resonance (MR) imaging at follow-up to further confirm tumor cell viability. If complete resolution occurred after repeated chemoembolization, patients were followed up by using dynamic CT scanning and laboratory tests (including measurements of serum α-fetoprotein concentration) at 2- or 3-month intervals until tumor recurrence. The median follow-up duration, defined as the time between baseline and death or censoring, was 24 months (interquartile range [IQR], 12.6–37.6 months; full range, 1.5–93.7 months).

Measurement of Treatment Response

The four anatomic imaging criteria are algorithmically defined in Figure 2. For one-dimensional measurement, we used the RECIST size and mRECIST enhancement criteria, respectively (Figs 3, 4, *A* and *B*). For the WHO size and EASL enhancement criteria, respectively, a binary method was used to calculate any response (Figs 3, 4, *C* and *D*). All criteria embraced the following four response categories: CR, PR, SD, and PD. Objective response included both CR and PR. According to RECIST, CR was defined as the absence of arterially enhanced areas in all target lesions; PR and PD, as a greater than 30% decrease and a greater than 20% increase, respectively, of the sum of the longest diameters of the enhancing target lesions; and SD, as neither PR nor PD (7). According to EASL criteria, PR and PD were defined as a greater than 50% decrease and a greater than 25% increase, respectively, of the sum of the cross products of the enhancing target lesions (2). The appearance of new HCC lesions denoted PD under both criteria, confirmed when their diameter exceeded 1 cm or when the lesion became at least 1 cm larger on progressive scans (7). Even if this finding was retrospectively confirmed with subsequent imaging, we defined the time of PD as the day when the new lesion was first detected with radiologic testing (6). Intratumoral iodized oil deposits were rated as necrotic areas by using enhancement criteria (9).

The outcome measurements of anatomic tumor burden and changes in such a burden were assessed by using the arterial-dominant phase of the dynamic CT scan at baseline and at follow-up,

Figure 2

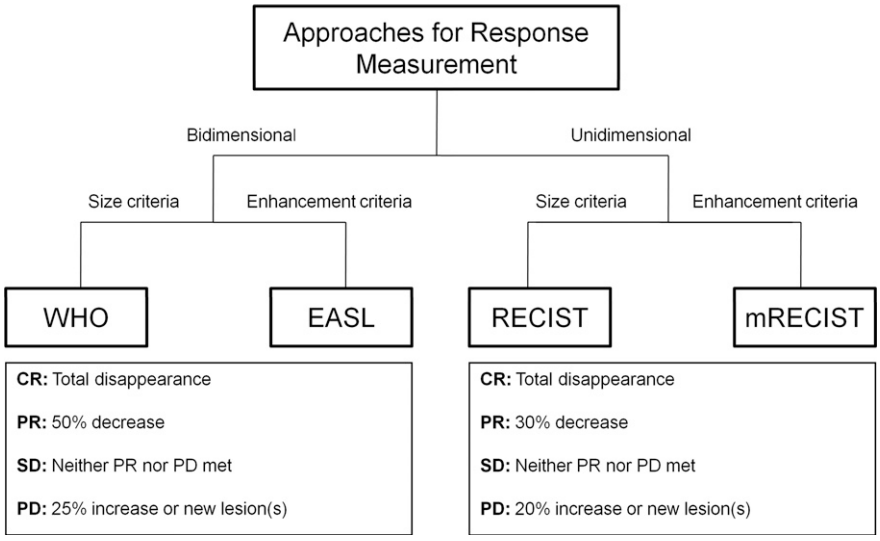

Figure 2: Radiographic response criteria used to assess the clinical effects of HCC treatment. CR = complete response, PD = progressive disease, PR = partial response, SD = stable disease.

Figure 3

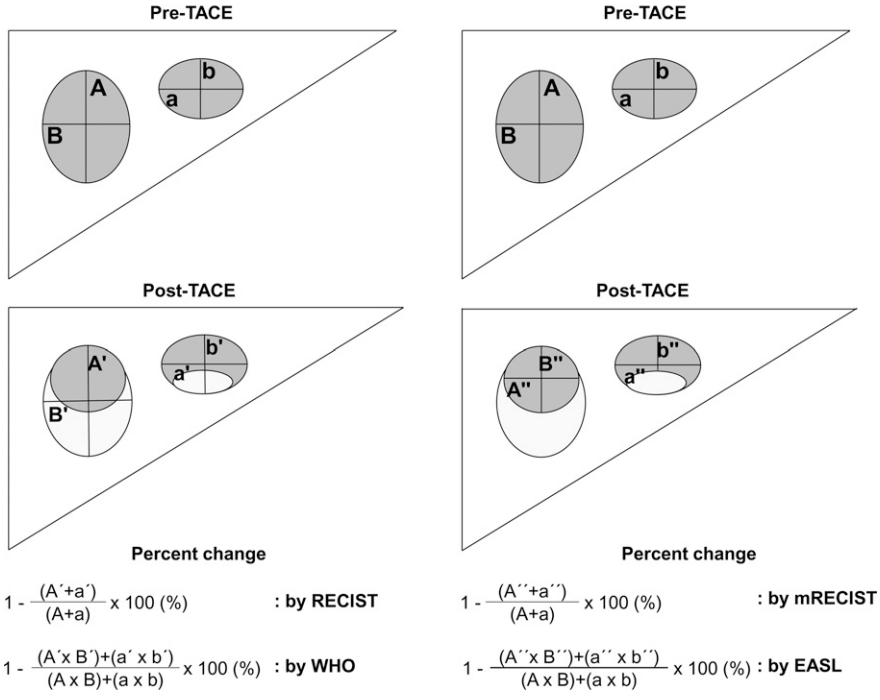

Figure 3: Illustration of the methods of measurement according to the four response criteria. The gray area within the nodules represents enhancing viable lesions, and the white area represents nonenhancing or iodized oil–retaining lesions. *A* = maximum diameter of entire larger tumor before TACE *(Pre-TACE)*; *a* = maximum diameter of entire smaller tumor before TACE; *A´* = maximum diameter of entire larger tumor after TACE *(Post-TACE)*; *a´* = maximum diameter of entire smaller tumor after TACE; *A´´* = maximum diameter of enhancing larger area of tumor after TACE; *a´´* = maximum diameter of enhancing area of smaller tumor after TACE; *B* = diameter perpendicular to A; *b* = diameter perpendicular to a; *B´* = diameter perpendicular to A´; *b´* = diameter perpendicular to a´; *B´´* = diameter perpendicular to A´´; *b´´* = diameter perpendicular to, a´´.

as this was the most accurate and reproducible manner by which we could assess tumor response (7,10–12). CT scanning was performed with 5-mm-thick contiguous sections and a multiphasic liver protocol by using spiral multidetector CT scanners (LightSpeed QX/I or LightSpeed Plus, GE Medical Systems, Milwaukee, Wis; or Somatom Sensation 16, Siemens, Erlangen, Germany), consistent with accepted guidelines (12). CT sections were acquired uniformly during the hepatic arterial, portal venous, and equilibrium phases at 36 seconds (or by using a bolus-tracking technique), 72 seconds, and 180 seconds, respectively, after the start of contrast agent infusion. In each patient, a total of 120–150 mL of iopromide (Ultravist 300 or Ultravist 370; Schering, Berlin, Germany) was intravenously administered at a rate of 3 mL/sec by using an automatic power injector (MCT Plus; Medrad, Pittsburgh, Pa).

Any measurable lesions with diameters 1 cm or greater at baseline were considered to be target lesions and were examined before and after treatment. Thus, for each target lesion, we evaluated a single diameter or the product of two dimensions of enhancing viable tissues. Table 1 illustrates the net response categories in accordance with possible combinations of tumor responses in target and nontarget lesions with or without the appearance of new malignant lesions (5,12). For example, all nontarget lesions must also have shown CR, without de novo lesions, to classify the final overall response as CR (Fig 5).

The best overall-response pattern of each patient, from commencement of the initial chemoembolization session to the end of a series of repeated sessions, was assigned to a final-response category (6). To minimize the possibility of false categorizations, all measurements were performed by an independent observer (Y.M.S) who was blinded to clinical data (13). Whenever response categorization was not obvious, final classification was made with consensus (Y.M.S., J.H.S., H.C.L., with 21, 10, and 22 years of experience, respectively).

Figure 4

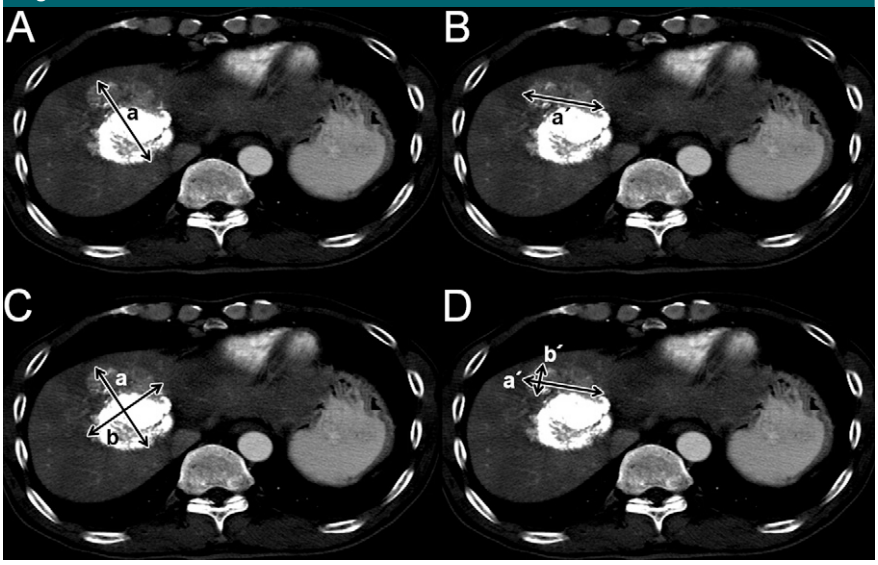

Figure 4: Axial CT scan obtained during arterial hepatic phase 1 month after initial TACE in 61-year-old man shows one HCC lesion treated with TACE. *A, B,* Unidimensional measurements. C, D, Bidimensional measurements. *A,* RECIST, *a*; *B,* mRECIST, *a´*; *C,* WHO, $a \times b$; and *D,* EASL criteria, $a' \times b'$ (a = maximum diameter of entire tumor, b = diameter perpendicular to a, a' = maximum diameter of enhancing tumor, b' = diameter perpendicular to a').

Figure 5

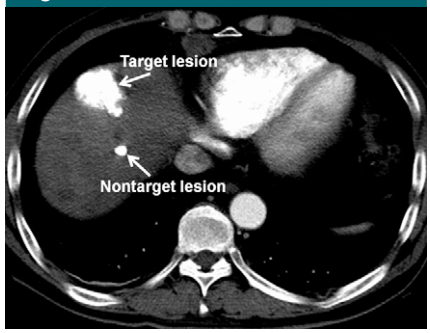

Figure 5: Axial CT scan obtained during hepatic arterial phase 1 month after initial TACE in 66-year-old man shows overall response of CR as a result of combined CR classifications for both target and nontarget lesions without the appearance of new lesions.

Table 1

Overall Responses Determined with Evaluation of Target, Nontarget, and New Lesions

Target Lesions	Nontarget Lesions	New Lesions	Overall Response
CR	CR	Absent	CR
CR	Non-CR or non-PD	Absent	PR
CR	Unmeasurable	Absent	PR
PR	Non-PD or unmeasurable	Absent	PR
SD	Non-PD or unmeasurable	Absent	SD
PD	Any response	Present or absent	PD
Any response	PD	Present or absent	PD
Any response	Any response	Present	PD

Statistical Analysis

Interassessment concordance between similar categorical items of the four criteria was measured with use of the κ coefficient. The strength of agreement based on κ values was interpreted as follows: κ less than 0.21, poor; κ of 0.21–0.40, fair; κ of 0.41–0.60, moderate; κ of 0.61–0.80, good; and κ greater than 0.80, excellent (14). We investigated the properties of the overall *C* index as a natural extension of the area under the receiver operating characteristic curve for survival analysis as a means of assessing discrimination with the four competing models (15). A *C* index of 0.5 or less indicates prediction no better than chance, and values from 0.5 to 1.0 (perfect prediction) indicate improvement over chance (16). Survival curves were estimated with the Kaplan-Meier method and were compared by using the log-rank test in accordance with the final-response outcomes. Survival time was evaluated from the day of initial treatment to the day of death, regardless of the cause of death; there were no procedure-related deaths within 1 month of the initial therapy (17). Survival calculations were censored at change of therapy following repeated chemoembolization. For calculation of the time to progression (TTP), radiologic progression as defined by the EASL and mRECIST guidelines was used, and deaths during follow-up without evidence of radiologic progression were censored. Two survival prediction models were constructed by using the Cox proportional hazards regression model, including either EASL or mRECIST guidelines as covariates. A multivariate Cox model was used to estimate mRECIST assessment of hazard increment, independent of other explanatory covariates. With a Cox regression model of the log hazard ratio (HR) on an objective response with a standard deviation of 0.346 that was based on a sample of 332 observations, a 95.2% power can be achieved at a .05 significance level for detecting a regression coefficient equal to −0.755 (HR, 0.47) (11). The sample size was adjusted for an anticipated event rate of 0.58.

Results

Baseline Demographics

Table 2 summarizes the baseline demographics of our cohort of 332 patients. Median patient age was 62 years (IQR, 55–68; full range, 34–82); most of the patients were men (87.3%); and most tested positive for hepatitis

Table 2

Baseline Characteristics of 332 Patients with BCLC Stage B HCC

Characteristics	No. of Patients*	Percentage of Patients
Age (y)		
Median	62 (55–68, 34–82)	. . .
≤60	133	40.1
>60	199	59.9
Sex		
M	290	87.3
F	42	12.7
Cause of liver disease		
Hepatitis B virus	255	76.8
Hepatitis C virus	42	12.7
Other	35	10.5
MELD score		
Median	8 (7–9,6–12)	. . .
≤8	214	64.5
>8	118	35.5
Basis of HCC diagnosis		
Radiologic evidence	255	76.8
Pathologic evidence	77	23.2
No. of tumors		
2–3	197	59.3
≥4	135	40.7
Maximum tumor diameter (cm)		
Median	4.7 (3.6–6.9; 1.6–15.9)	. . .
≤5	181	54.5
>5	151	45.5
Serum α-fetoprotein level (ng/mL)†		
Median	85.2 (13.8–1090; 1.0–477 000)	. . .
≤200	201	60.5
>200	131	39.5

Note.—MELD = model for end-stage liver disease.

* Numbers in parentheses are the IQRs and ranges, respectively.

† To convert to Système International units in micrograms per liter, multiply by 1.0.

B virus (76.8%). Their median model for end-stage liver disease score was eight (IQR, 7–9; full range, 6–12). As all patients had the typical features of HCC at CT, in only 77 (23.2%) of 332 patients with HCC was a pathologic diagnosis determined on the basis of American Association for the Study of Liver Diseases guidelines (1). At initial chemoembolization, more than one-half of the patients (59.3%) had fewer than four tumors; the median diameter of the largest nodule was 4.7 cm (IQR, 3.6–6.9 cm; full range, 1.6–15.9 cm), and 45.5% of patients had lesions larger than 5 cm. Of all patients, 39.5% had α-fetoprotein levels greater than 200 ng/mL (200 μg/L) before TACE.

Intercriterion Agreement

The median number of target lesions evaluated per patient was two (range, 1–19), and we measured at least three target lesions in 25% ($n = 83$) of patients. When any measurement guideline was evaluated, a median of two TACE sessions (range, 1–6) was performed prior to assessment of the best response. Intercriterion agreement evaluation by using the Cohen κ statistic showed that, in spite of good correlation between each other (κ = 0.80; 95% confidence interval [CI]: 0.74, 0.86) (Table 3), data of the WHO and original RECIST models correlated poorly with regard to both enhancement criteria (κ < 0.20 for both comparisons) (Table 4). However, we noted excellent agreement between EASL and mRECIST guidelines, as reflected by a κ value of 0.94 (95% CI: 0.91, 0.97) (Table 3).

Survival Stratification Models according to the Four Response Criteria

Figure 6 shows the Kaplan-Meier survival curves of HCC patients treated with serial chemoembolization, by using the criteria of the four response measurement protocols. A total of 81 patients were censored when their treatment was changed following repeated chemoembolization; the changes involved radiofrequency ablation in 62 patients, percutaneous ethanol injection in four, and surgical resection in 15. Response category curves contained four visible peaks in the mRECIST model, whereas curves crossed or were very close to each other over time in the other models. The mRECIST model yielded a significant difference in the probability of survival across the different response categories in the model, when analyzed with the log-rank test ($P < .001$ for each comparison). On the other hand, there was no significant difference in the probability of survival between PR and SD categories with the EASL model ($P = .71$).

Comparison of End Points according to Degree of Response and Enhancement Criteria

For all 332 patients, median TTP and OS were 15 months (95% CI: 13.1, 16.9) and 34.3 months (95% CI: 31.3, 37.3), respectively. With mRECIST and EASL criteria, we identified 278 (83.7%) and 286 (86.1%) responders, respectively. When the mRECIST definition of an objective response (ie, CR or PR) was used, HRs for TTP and OS in responders compared with nonresponders were 0.15 (95% CI: 0.10, 0.21) and 0.18 (95% CI: 0.13, 0.26), respectively ($P < .001$ for both) (Table 6). With the EASL criteria, the differences in TTP and OS between responders and nonresponders remained significant (0.16 [95% CI: 0.11, 0.22] and 0.23 [95% CI: 0.16, 0.33], respectively; $P < .001$ for both) (Table 6).

While patients with PD were similar when we used all four assessment

Table 3

Intercriterion Agreement between the Size Criteria and between the Enhancement Criteria

A: Agreement between WHO and RECIST Criteria

	Response with RECIST Criteria			
Response with WHO Criteria	CR	PR	SD	PD
CR	0	0	0	0
PR	0	113	28	0
SD	0	8	151	0
PD	0	0	2	30

B: Agreement between EASL and mRECIST Criteria

	Response with mRECIST Criteria			
Response with EASL Criteria	CR	PR	SD	PD
CR	135	0	0	0
PR	0	141	10	0
SD	0	2	14	0
PD	0	0	0	30

Note.—Data are the numbers of patients. The κ value for agreement between WHO and RECIST criteria was 0.80 (95% CI: 0.74, 0.86). The κ value for agreement between EASL and mRECIST criteria was 0.94 (95% CI: 0.91, 0.97).

Table 4

Intercriterion Agreement between Size and Enhancement Estimations

A: Agreement between WHO and EASL and WHO and mRECIST Criteria

	Response with EASL Criteria				Response with mRECIST Criteria			
Response with WHO Criteria	CR	PR	SD	PD	CR	PR	SD	PD
CR	0	0	0	0	0	0	0	0
PR	68	73	0	0	68	72	1	0
SD	67	76	16	0	67	70	22	0
PD	0	2	0	30	0	1	1	30

B: Agreement between RECIST and EASL and RECIST and mRECIST Criteria

	Response with EASL Criteria				Response with mRECIST Criteria			
Response with RECIST Criteria	CR	PR	SD	PD	CR	PR	SD	PD
CR	0	0	0	0	0	0	0	0
PR	57	63	1	0	57	64	0	0
SD	78	88	15	0	78	79	24	0
PD	0	0	0	30	0	0	0	30

Note.—Data are the numbers of patients. The κ value for agreement between WHO and EASL criteria was 0.17 (95% CI: 0.12, 0.23) and that for agreement between WHO and mRECIST criteria was 0.19 (95% CI: 0.13, 0.25). The κ value for agreement between RECIST and EASL criteria was 0.16 (95% CI: 0.10, 0.21) and that for agreement between RECIST and mRECIST was 0.19 (95% CI: 0.14, 0.24).

guidelines, because 90% of those with PD had new HCC lesions appearing after the first chemoembolization, patients with CR were quantitatively the same when assessed with EASL or mRECIST guidelines but were not identified by using uni- or bidimensional tumor size measurements (Tables 3, 4, 6). The HRs for TTP and OS of patients with a CR were 0.52 (95% CI: 0.40, 0.67) and 0.29 (95% CI: 0.21, 0.40), respectively ($P <$ .001 for both) (Table 6).

Outcome Prediction on the Basis of Enhancement Criteria

Table 5 presents the results of the models for predicting TTP and OS constructed by using the Cox regression model, on the basis of the EASL and mRECIST guidelines. The HR levels for TTP and OS compared with CR were escalated in a stepwise manner as the response progressed toward PD in both EASL and mRECIST models. The *C* indexes of the survival prediction models were 0.71 (95% CI: 0.67, 0.74) for EASL and 0.72 (95% CI: 0.68, 0.76) for mRECIST, greater than 0.67 (95% CI: 0.64, 0.71) for WHO and RECIST, thus indicating the models with the greater capability to predict survival (Table 5).

Survival of Complete Responders according to Enhancement and Size Criteria

Among a subgroup of 135 patients with a CR, as determined with enhancement criteria, 68 and 67 patients were considered to show PR and SD, respectively, according to WHO guidelines, and 57 and 78 patients showed PR and SD, respectively, according to RECIST guidelines (Table 4). Of the latter patients, no significant bifurcation in PR and SD survival curves was obvious with respect to original size criteria (Fig 7).

Survival Prediction on the Basis of Enhancement Criteria, Adjusted by Covariates

A multivariate Cox model exploring prognostic factors for OS is shown in Table 7. The HRs were adjusted for age, sex, hepatitis B virus infection status, α-fetoprotein level, model for end-stage liver disease score, tumor size, and EASL or mRECIST guidelines, by using the cutoff points shown in Table 2. After adjustment, multivariate analysis revealed a significant independence of the mRECIST model in predicting OS in patients with HCC treated with chemoembolization, showing escalation of HR levels as the response progressed toward PD (adjusted HRs, 3.05 [95% CI: 2.08, 4.47], 6.42 [95% CI: 3.69, 11.15], and 12.14 [95% CI: 7.04, 20.93]; $P <$ .001). Similar results were also obtained with the EASL model (adjusted HRs, 3.48 [95% CI: 2.40, 5.05], 3.39 [95% CI: 1.64, 6.98], and 12.09 [95% CI: 7.01, 20.85]; $P <$ .001). The *C* index of the multivariate prediction models was 0.76 (95% CI: 0.72, 0.79), which was the same for the two enhancement criteria (Table 7).

Figure 6

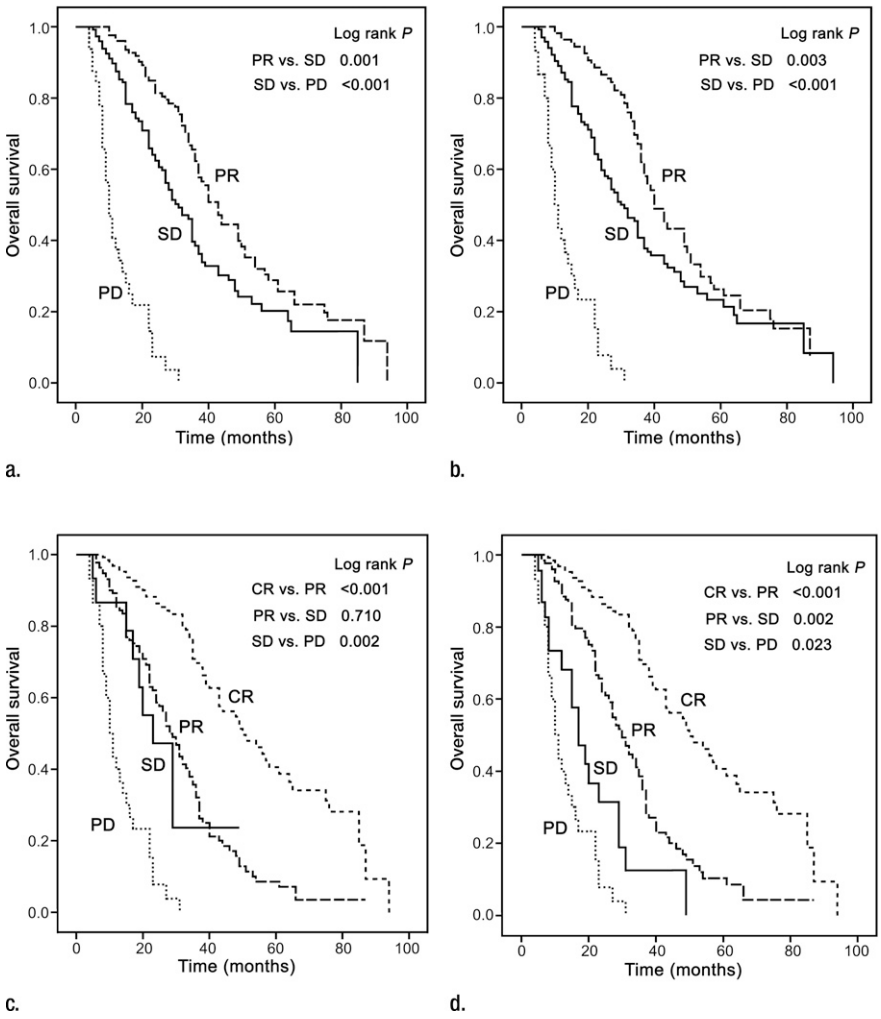

Figure 6: Survival of 332 BCLC stage B patients, as determined with the **(a)** WHO, **(b)** RECIST, **(c)** EASL, and **(d)** mRECIST guidelines. Data were stratified into four response categories. The HRs for PR, SD, and PD, compared with CR, by using the EASL and mRECIST models, are described in Table 5.

Discussion

In 2008, a complementary framework to assess therapeutic response was formally introduced on the basis of guidelines established by the American Association for the Study of Liver Diseases–Journal of the National Cancer Institute. The revised guidelines took into account the concept of tumor viability (ie, from EASL criteria) and a single linear summation (ie, from RECIST data) (7). We found that this mRECIST assessment showed excellent intercriterion agreement with the EASL data and also a clear correlation between response rate and survival prediction, in line with the EASL assessment. In particular, our data suggest that the radiologic enhancement guidelines for HCC are more reliable measures of tumor response to chemoembolization, on the basis of the capability to predict survival across various patient categories; this was supported by the data for the *C* statistic. These findings corresponded with the TTP data.

The OS and TTP are generally accepted to be the major end points for clinical trials in HCC (2,6). However, rather long follow-up times are required to assess OS and TTP, especially in patients with early-stage HCC and good liver function, and it is important, clinically, to determine whether current interventions should be continued or altered. Although response rate can be relatively quickly assessed, this is a weaker end point, because accuracy depends on investigator skill and appropriateness of the reference criteria (6,7). In such a practical context, simple and time-saving categorical response guidelines reflecting survival outcomes, and specific to local-regional therapy, must be standardized. The EASL panel of HCC experts suggested new imaging criteria that could be used to evaluate histologic tumor necrotic reaction following treatment. Such responses appear as arterially nonenhanced areas within treated lesions at dynamic CT or MR imaging (2). Researchers in recent studies have reported that the EASL criteria provide a significant anatomic response and are useful prognosticators of clinical outcomes in HCC patients following chemoembolization or yttrium 90 radioembolization (11,18).

With the EASL guidelines, a bidimensional method is used to estimate tumor load, on the basis of WHO size criteria. However, WHO criteria are less frequently used in anticancer therapeutic trials. Rather, in such trials, original RECIST assessment is used, which reduces interobserver variability, and is comprehensive, consistent, and timely, particularly when newer imaging technologies are used (12,19). More important, approaches with which cross products are used would contain inherent physical and mathematic limitations (20). The simplicity of the calculation encourages the measurement of more lesions in an individual patient, which may reduce false categorization of responses (13,20). These considerations may lead to favoring the use of mRECIST evaluation.

Our findings clearly reveal a high degree of discordance between size and enhancement criteria in evaluation of HCC progression, consistent with findings in previous reports (10,11). Patients in the CR category were reflected in both enhancement data and those reflecting improved survival. Notably,

Table 5

Predictive Models That Are Based on the Enhancement Criteria

	TTP		OS	
Predictive Response	HR*	*P* Value†	HR*	*P* Value†
With EASL criteria				
CR	1.0	. . .	1.0	. . .
PR	1.57 (1.19, 2.06)	.001	2.99 (2.14, 4.17)	<.001
SD	2.89 (1.54, 5.44)	.001	3.49 (1.71, 7.10)	.001
PD	49.89 (29.18, 85.30)	<.001	15.63 (9.51, 25.69)	<.001
With mRECIST criteria				
CR	1.0	. . .	1.0	. . .
PR	1.52 (1.16, 2.01)	.003	2.75 (1.96, 3.87)	<.001
SD	3.75 (2.14, 6.57)	<.001	6.32 (3.67, 10.90)	<.001
PD	52.30 (30.49, 89.70)	<.001	16.06 (9.76, 26.43)	<.001

Note.—The *C* index for EASL criteria was 0.71 (95% CI: 0.67, 0.74), and that for mRECIST criteria was 0.72 (95% CI: 0.68, 0.76).

* Numbers in parentheses are the 95% CIs.

† Data were generated from the univariate Cox regression model.

Table 6

End Point Prediction according to the Degree of Response

Parameter	No. of Patients*	TTP HR†	OS HR†
EASL criteria			
Responders	286 (86.1)	0.16 (0.11, 0.22)	0.23 (0.16, 0.33)
Nonresponders	46 (13.9)	1.0	1.0
Complete responders	135 (40.7)	0.52 (0.40, 0.67)	0.29 (0.21, 0.40)
Noncomplete responders	197 (59.3)	1.0	1.0
mRECIST criteria			
Responders	278 (83.7)	0.15 (0.10, 0.21)	0.18 (0.13, 0.26)
Nonresponders	54 (16.3)	1.0	1.0
Complete responders	135 (40.7)	0.52 (0.40, 0.67)	0.29 (0.21, 0.40)
Noncomplete responders	197 (59.3)	1.0	1.0

Note.—The *P* value for all comparisons was less than .001, and the data were generated from the univariate Cox regression model.

* Numbers in parentheses are percentages. Percentages were rounded.

† Numbers in parentheses are the 95% CIs.

among a subset of patients experiencing CR according to enhancement guidelines, patients with PR paralleled those with SD in terms of survival pattern when size criteria were used. This finding revealed the inappropriateness of guidelines that were based solely on tumor size, as well as emphasized the importance of internal necrosis.

In a recent phase II study of brivanib (Bristol-Myers Squibb, Princeton, NJ) treatment in patients with HCC, mRECIST criteria showed a greater alignment of TTP with a clearer identification of PD than modified WHO criteria (21). In the present study, however, outcomes of patients with PD were similar with respect to all four assessment guidelines, and these patients experienced the worst survival outcomes. Responses in 27 (90%) of 30 patients with PD were unequivocally defined by the appearance of new lesions after the first chemoembolization, rather than by measurement-based progression. These findings may also explain why a shorter time to best response unexpectedly correlated with poor patient survival (Table 7).

Our study sought to control for measurement bias and variability: *(a)* We selected easily measurable index nodules 1 cm or larger in diameter as target lesions, to preclude error. *(b)* We included all target lesions in calculation and summation to represent the overall tumor response. *(c)* We used the same CT technique to characterize each identifiable lesion at baseline and at follow-up examinations for each patient (12).

In the present study, we used a newer method to validate four response assessment models, employing survival data in which two important predictors of survival (HCC stage and underlying liver function) were equivalent in all subjects. Researchers in studies seeking to validate enhancement criteria included a group of HCC patients treated with different local-regional modalities who had various levels of hepatic function and tumor status, although these variables were treated as possible confounders of survival in a multivariate analysis (10,11). In contrast, our method substantially eliminates confounding relationships between response categories and outcome predictions.

A potential weakness of our work is the lack of cytopathologic confirmation of radiologic measurements. However, pathologic explants cannot reveal the effects of treatment before explantation on current survival times. The time lag bias that results from the interval between treatment and pathologic evaluation may prevent direct correlation of these data. Although radiologic nonenhancement may not allow complete differentiation of viable from histopathologic necrotic tumors (22), the use of imaging-based criteria measuring areas of intratumoral enhancement is of great clinical value in the estimation of treatment response, particularly in instances where the information obtained can be used to predict survival. Another limitation is that, for evaluating responses, we did not always use a dynamic MR imaging examination, which may be better than dynamic CT for detecting small nodules in patients treated with chemoembolization (23). Further validation with an MR imaging system by using a currently available liver-specific

Figure 7

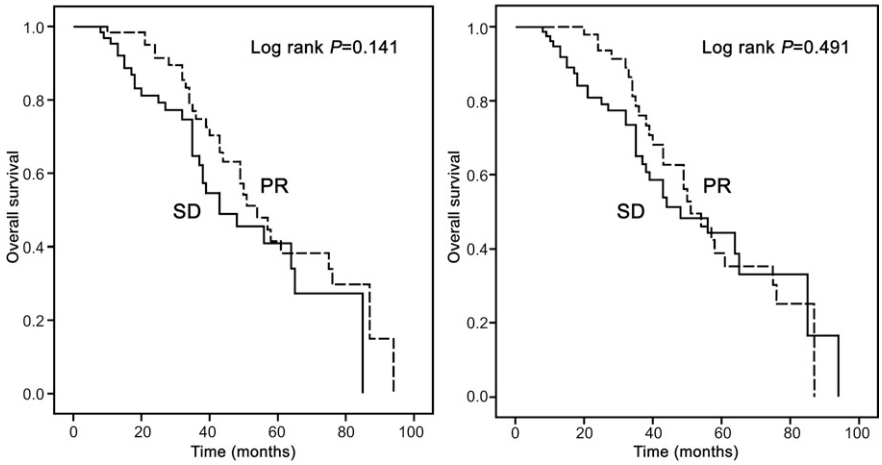

Figure 7: Survival of patients with PR and SD according to the **(a)** WHO and **(b)** RECIST models for patients in whom a CR, as determined by using the enhancement guidelines, was achieved. Among the subgroup of patients showing a CR, as determined by using the EASL or mRECIST models, those with a PR were similar to patients with SD in terms of median survival when either the WHO or RECIST criteria were used for evaluation. Median survival was 54.0 months (95% CI: 43.9 months, 64.1 months) versus 43.0 months (95% CI: 26.4 months, 59.6 months) when WHO criteria were used ($P = .141$) and 51.0 months (95% CI: 41.3 months, 60.9 months) versus 48.0 months (95% CI: 31.1 months, 64.9 months) when RECIST criteria were used ($P = .491$).

Table 7

Accuracy of Prediction of Patient Survival from Assessment of Responses

A: Accuracy in Terms of Clinical Data

Parameter	Adjusted HR*	P Value†	Adjusted HR*	*P* Value†
α-Fetoprotein level > 200 ng/mL‡	1.49 (1.11, 2.00)	.008	1.56 (1.16, 2.10)	.003
No. of tumors ≥ 4	1.39 (1.01, 1.92)	.042	1.45 (1.05, 1.99)	.025
Time to best response (per month)	0.92 (0.88, 0.97)	<.001	0.93 (0.89, 0.98)	.003

B: Accuracy in Terms of Response and Criteria

Response Category	EASL Criteria HR*	*P* Value†	mRECIST Criteria HR*	*P* Value†
CR	1.0	. . .	1.0	. . .
PR	3.48 (2.40, 5.05)	<.001	3.05 (2.08, 4.47)	<.001
SD	3.39 (1.64, 6.98)	.001	6.42 (3.69, 11.15)	<.001
PD	12.09 (7.01, 20.85)	<.001	12.14 (7.04, 20.93)	<.001

Note.—The *C* index for response category for EASL criteria and for mRECIST criteria was 0.76 (95% CI: 0.72, 0.79), with $P < .001$.

* Numbers in parentheses are the 95% CIs.

† Data were generated from the multivariate Cox regression model. Covariates with $P < .2$ from the univariate analysis were then included in the multivariate analysis.

‡ To convert to Système International units in micrograms per liter, multiply by 1.0.

contrast agent may be required to draw concrete conclusions on the usefulness of the enhancement criteria.

Our results suggest that enhancement approaches, rather than any size criteria, are better anatomic criteria for categorizing distinct responses in trials assessing the survival or TTP of patients with HCC following chemoembolization; in these trials, one-dimensional mRECIST measurement may simplify manual efforts and, thus, reduce workload. The results of our analysis warrant further studies to determine whether the enhancement model can be extrapolated to responses to percutaneous or other transarterial interventions.

Disclosures of Potential Conflicts of Interest: J.H.S. No potential conflicts of interest to disclose. **H.C.L.** No potential conflicts of interest to disclose. **S.O.K.** No potential conflicts of interest to disclose. **Y.M.S.** No potential conflicts of interest to disclose. **K.M.K.** No potential conflicts of interest to disclose. **Y.S.L.** No potential conflicts of interest to disclose. **D.J.S.** No potential conflicts of interest to disclose.

References

1. Bruix J, Sherman M; American Association for the Study of Liver Diseases. Management of hepatocellular carcinoma: an update. Hepatology 2011;53(3):1020–1022.
2. Bruix J, Sherman M, Llovet JM, et al. Clinical management of hepatocellular carcinoma: conclusions of the Barcelona-2000 EASL conference. European Association for the Study of the Liver. J Hepatol 2001;35(3):421–430.
3. Llovet JM, Real MI, Montaña 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.
4. Miller AB, Hoogstraten B, Staquet M, Winkler A. Reporting results of cancer treatment. Cancer 1981;47(1):207–214.
5. Therasse P, Arbuck SG, Eisenhauer EA, et al. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. J Natl Cancer Inst 2000;92(3): 205–216.
6. Llovet JM, Di Bisceglie AM, Bruix J, et al. Design and endpoints of clinical trials in hepatocellular carcinoma. J Natl Cancer Inst 2008;100(10):698–711.
7. Lencioni R, Llovet JM. Modified RECIST (mRECIST) assessment for hepatocellular carcinoma. Semin Liver Dis 2010;30(1): 52–60.
8. Kasugai H, Kojima J, Tatsuta M, et al. Treatment of hepatocellular carcinoma by transcatheter arterial embolization combined with intraarterial infusion of a mixture of cisplatin and ethiodized oil. Gastroenterology 1989;97(4):965–971.

9. Choi BI, Kim HC, Han JK, et al. Therapeutic effect of transcatheter oily chemoembolization therapy for encapsulated nodular hepatocellular carcinoma: CT and pathologic findings. Radiology 1992;182(3):709–713.

10. Forner A, Ayuso C, Varela M, et al. Evaluation of tumor response after locoregional therapies in hepatocellular carcinoma: are response evaluation criteria in solid tumors reliable? Cancer 2009;115(3):616–623.

11. Riaz A, Miller FH, Kulik LM, et al. Imaging response in the primary index lesion and clinical outcomes following transarterial locoregional therapy for hepatocellular carcinoma. JAMA 2010;303(11):1062–1069.

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

13. Warr D, McKinney S, Tannock I. Influence of measurement error on assessment of response to anticancer chemotherapy: proposal for new criteria of tumor response. J Clin Oncol 1984;2(9):1040–1046.

14. Altman D. Practical statistics for medical research. London, England: Chapman & Hall, 1991.

15. Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. Stat Med 2004;23(13):2109–2123.

16. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143(1):29–36.

17. Takayasu K, Arii S, Ikai I, et al. Prospective cohort study of transarterial chemoembolization for unresectable hepatocellular carcinoma in 8510 patients. Gastroenterology 2006;131(2):461–469.

18. Duke E, Deng J, Ibrahim SM, et al. Agreement between competing imaging measures of response of hepatocellular carcinoma to yttrium-90 radioembolization. J Vasc Interv Radiol 2010;21(4):515–521.

19. Mazumdar M, Smith A, Schwartz LH. A statistical simulation study finds discordance between WHO criteria and RECIST guideline. J Clin Epidemiol 2004;57(4):358–365.

20. James K, Eisenhauer E, Christian M, et al. Measuring response in solid tumors: unidimensional versus bidimensional measurement. J Natl Cancer Inst 1999;91(6):523–528.

21. Raoul JL, Lencioni R, Park JW, Baudelet C, Walters I. Tumor response in a phase 2 study of first- and second-line brivanib in hepatocellular carcinoma (HCC): comparison of modified WHO and modified RECIST criteria [abstr]. Presented at the 45th annual meeting of the European Association for the Study of the Liver, Vienna, Austria, April 14–18, 2010.

22. Maddala YK, Stadheim L, Andrews JC, et al. Drop-out rates of patients with hepatocellular cancer listed for liver transplantation: outcome with chemoembolization. Liver Transpl 2004;10(3):449–455.

23. Kim T, Murakami T, Oi H, et al. Detection of hypervascular hepatocellular carcinoma by dynamic MRI and dynamic spiral CT. J Comput Assist Tomogr 1995;19(6):948–954.