Clinical Cancer Research

$\mathcal{AACR}$

Number of Target Lesions for EASL and Modified RECIST to Predict Survivals in Hepatocellular Carcinoma Treated with Chemoembolization

Beom Kyung Kim, Seung Up Kim, Myeong-Jin Kim, et al.

Updated version	Access the most recent version of this article at: doi:10.1158/1078-0432.CCR-12-2721
Supplementary Material	Access the most recent supplemental material at: http://clincancerres.aacrjournals.org/content/suppl/2012/12/07/1078-0432.CCR-12-2721.DC1.html http://clincancerres.aacrjournals.org/content/suppl/2013/03/20/1078-0432.CCR-12-2721.DC2.html

Cited Articles	This article cites by 30 articles, 6 of which you can access for free at: http://clincancerres.aacrjournals.org/content/19/6/1503.full.html#ref-list-1
Citing articles	This article has been cited by 1 HighWire-hosted articles. Access the articles at: http://clincancerres.aacrjournals.org/content/19/6/1503.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.

Imaging, Diagnosis, Prognosis
  See related commentary by Lencioni, p. 1312

Number of Target Lesions for EASL and Modified RECIST to Predict Survivals in Hepatocellular Carcinoma Treated with Chemoembolization

Beom Kyung Kim[1,4], Seung Up Kim[1,3,4], Myeong-Jin Kim[2], Kyung Ah Kim[2], Do Young Kim[1,3,4], Jun Yong Park[1,3,4], Sang Hoon Ahn[1,3,4,5], Kwang-Hyub Han[1,3,4,5], and Chae Yoon Chon[1,3,4]

Abstract

Purposes: To date, most studies about the optimal number of target lesions for enhancement criteria for hepatocellular carcinoma (HCC) have focused on cross-sectional analyses of concordance. We investigated the optimal number of target lesions for European Association for the Study of the Liver (EASL) and modified Response Evaluation Criteria in Solid Tumors (mRECIST) guidelines in predicting overall survival (OS).

Experimental Design: We analyzed 254 consecutive treatment-naïve patients with HCC having at least 2 measurable target lesions undergoing transarterial chemoembolization. Kappa values for intermethod agreement of treatment responses were calculated for comparisons between use of maximum of 1, 2, 3, 4, or 5 targets versus use of all target lesions. Prognostic values of radiologic assessments according to number of target lesions for predicting OS were expressed as C-index.

Results: By EASL and mRECIST guidelines, κ values between responses assessing the longest 2, 3, 4, or 5 targets and assessing all targets were 0.924, 0.977, 1.000, or 1.000 and 0.907, 0.959, 1.000, or 1.000, respectively, whereas those between responses assessing only one target and assessing all target lesions were 0.723 and 0.666, respectively. C-index when measuring the longest 1, 2, 3, 4, 5, and all targets was similar, ranging from 0.739 to 0.749 for EASL criteria and from 0.750 to 0.759 for mRECIST. From Cox regression analyses, radiologic response from each calculation method showed independently significant effects on OS for both guidelines, regardless of number of target lesions.

Conclusions: Prognostic values for predicting OS were similar regardless of number of target lesions. Assessing the 2 largest targets rather than only 1 index lesion could be recommended considering high concordances from cross-sectional analyses. *Clin Cancer Res; 19(6); 1503–11. ©2012 AACR.*

Introduction

World Health Organization (WHO) and Response Evaluation Criteria in Solid Tumors (RECIST) guidelines have been universally accepted to evaluate treatment responses of solid tumors (1, 2). However, these 2 conventional size-based criteria designed primarily for evaluation of cytotoxic agents do not address measures of antitumor activity other than tumor shrinkage (2). In particular, for hepatocellular carcinoma (HCC), recent studies have shown poor correlations between conventional methods of response evalua-

tion and clinical benefits provided by molecular-targeted agents, transarterial chemoembolization (TACE), or other local ablative therapy, as these criteria generally ignore tumor necrosis or decreased tumor viability induced from such treatments (3–5). Furthermore, because regenerative nodules can appear as new lesions and tumors do not change in appearance on imaging due to preexisting fibrous matrix despite successful treatment, sized-based criteria leave much to be desired.

Therefore, the HCC panel of the European Association for the Study of the Liver (EASL) established consensus criteria in 2001, called EASL criteria, in which arterially enhanced tumor burden indicating the remaining viable tumor after treatment is calculated bidimensionally (6). Thereafter, a complementary framework to assess therapeutic response was formally introduced in 2008 based on guidelines established by the American Association for the Study of Liver Diseases—Journal of the National Cancer Institute (7). The revised guidelines, called modified RECIST (mRECIST; ref. 5), consider both concept of tumor viability based on arterial enhancement (i.e., from EASL criteria) and single linear summation (i.e., from RECIST; refs. 5, 7). In addition, mRECIST is a major step forward compared with the

Authors' Affiliations: Departments of [1]Internal Medicine and [2]Radiology; [3]Yonsei Institute of Gastroenterology, Yonsei University College of Medicine; [4]Liver Cirrhosis Clinical Research Center; and [5]Brain Korea 21 Project for Medical Science, Seoul, Republic of Korea

Corresponding Author: Seung Up Kim, Department of Internal Medicine, Yonsei University College of Medicine, 250 Seongsanno, Sodaemun-gu, Seoul 120-752, Republic of Korea. Phone: 82-2-2228-1982; Fax: 82-2-393-6884; E-mail: ksukorea@yuhs.ac

American Association for Cancer Research **1503**

Translational Relevance

Regarding European Association for the Study of the Liver (EASL) and modified Response Evaluation Criteria in Solid Tumors (mRECIST) guidelines for hepatocellular carcinoma (HCC), the least number of target lesions that should be measured to achieve an equal predictive value for long-term survival outcomes remains uncertain. The most ideal approach is to consider all target lesions; however, it requires much unnecessary time and labor. By EASL and mRECIST guidelines, C-index, which was calculated to evaluate prognostic values of each radiologic method according to number of target lesions (assessing the longest 1, 2, 3, 4, 5, and all targets) for predicting overall survival (OS), was similar, ranging from 0.739 to 0.749 for EASL criteria and from 0.750 to 0.759 for mRECIST, regardless of number of target lesions. In conclusion, prognostic values for predicting OS were similar regardless of number of target lesions.

previous enhancement method, EASL criteria, in terms that mRECIST not only somewhat simplifies complex EASL criteria, but also provides special recommendations for new lesion and nontarget lesions, such as portal vein thrombosis, lymph node at porta hepatis, ascites, or pleural effusion in detail. EASL and mRECIST guideline have shown superior efficacy for assessing treatment responses and predicting survival outcomes compared with WHO and RECIST guidelines in patients with HCC because enhancement criteria can discriminate patients with better clinical outcomes by tumor necrosis, regardless of shrinkage of entire tumor mass (3, 8–12).

Meanwhile, Warr and colleagues (13) showed that at least 2 or 3 index lesions should be measured to minimize false categorization of the final response for WHO criteria. And, the current RECIST criteria are designed to assess a maximum of 2 measurable target lesions per organ that are representative of all lesions within each organ (14). However, in contrast to size criteria, when using enhancement criteria, the least number of target lesions that should be measured to achieve an equal predictive value for long-term survival outcomes remains uncertain compared with what would be achieved if all lesions were considered. In particular, clinical outcomes about overall survival (OS), the robust and unequivocal endpoint for clinical trials, according to maximum number of target lesions have never been analyzed for enhancement criteria.

Here, we aimed to determine the optimal number of target lesions for EASL and mRECIST guidelines from perspectives of predicting survival outcomes among treatment-naïve patients with HCC undergoing TACE.

Materials and Methods

Patient eligibility

Through the retrospective review from the prospectively registered data bank from Yonsei Liver Cancer Special Clinic (Seoul, Republic of Korea), treatment-naïve patients with multifocal intrahepatic HCC who received first-line therapy with TACE between June 2006 and December 2009 were eligible for this study. On recruitment, exclusion criteria were as follows: presence of a solitary target lesion, inadequate target lesion (i.e., infiltrative pattern or largest lesion less than 1 cm), presence of an additional primary malignancy in another organ, presence of extrahepatic lesions or vascular invasion, Child–Pugh class B or C, and presence of uncontrolled functional or metabolic disease (Supplementary Fig. S1).

This study protocol was conducted in accordance with the ethical guidelines of the 1975 Declaration of Helsinki, and written informed consent was obtained from each participant or responsible family member. This study procedure was approved by the Institutional Review Board of Severance Hospital, Yonsei University College of Medicine (Seoul, Republic of Korea).

Diagnosis of HCC

Diagnosis of HCC and assessment of treatment response were conducted with a dynamic imaging study involving 4 phases (precontrast, arterial, portal, and equilibrium phases) using contrast-enhanced computed tomography (CT) or gadolinium-enhanced MRI as appropriate (5, 6, 14). Diagnosis of HCC was made based on guidelines proposed by Korea Liver Cancer Study Group (15). According to these criteria, a patient is considered positive for HCC if they have 1 or more risk factors (hepatitis B or C virus infection, cirrhosis) and one of the following: serum α-fetoprotein (AFP) more than 400 ng/mL and a positive finding on at least 1 of 3 typical imaging studies (dynamic CT, dynamic MRI, or hepatic angiography), or serum AFP of less than 400 ng/mL and positive findings on at least 2 of 3 imaging studies. A positive finding for typical HCC on dynamic CT or MRI was defined as increased arterial enhancement followed by decreased enhancement compared with liver (washout) in the portal or equilibrium phase.

Treatment modality

TACE was conducted by infusion with a mixture of 5 mL iodized oil contrast medium (lipiodol; Guerbet) and 50 mg adriamycin (Ildong Pharmaceutical) followed by embolization of feeding arteries using gelatin sponge particles (Cutanplast; Mascia Bruneili S.p.A.). Sequential TACE was scheduled at 6- to 8-week intervals when a residual viable tumor was detected in liver on a follow-up assessment without appearance of extrahepatic metastases, major portal vein invasion, or deterioration in clinical status or laboratory values.

Assessment of treatment responses using EASL and mRECIST guidelines

Both guidelines define viable tumors according to uptake of contrast material in the arterial phase of dynamic CT or MRI; tumors retaining iodized oil and necrotic lesions without intratumoral arterial enhancement were regarded

as necrotized tumor foci. Treatment responses were assessed 4 weeks after the initial TACE using both guidelines. EASL criteria are based on product of bidimensional diameters of enhanced area of measurable lesions, whereas mRECIST are based on sum of unidimensional measurements. All target lesions must be at least 10 mm in diameter and distinctly nodular (5).

Tumor response was quantitatively defined as complete response (CR), as indicated by complete disappearance of measurable lesions for both guidelines, or partial response (PR), defined as a 50% decrease from baseline for EASL criteria and a 30% decrease from baseline for mRECIST. Progressive disease was defined by a 25% increase from baseline for EASL criteria and a 20% increase from baseline for mRECIST. Stable disease was defined as a value between progressive disease and PR. Objective response rate referred to sum of CR and PR.

In per-patient agreement analysis, a response for a given patient based on all measurable lesions was computed and used as reference to further investigate potential outcome differences between assessment of 1, 2, 3, 4, or 5 target lesions. Target lesions were selected in order of their maximum diameter at baseline so as to be presumably representative of entire tumor burden: the longest 1, 2, 3, 4, and 5 lesions. Lesions other than designated target lesions for each methodology, always including any small lesions with a maximum diameter of less than 10 mm and truly nonmeasurable lesions, were considered as nontarget lesions.

For both guidelines, target responses take into account changes in only designated target lesions, whereas overall responses comprehensively take into consideration changes in target lesions, nontarget lesions, and appearance of intrahepatic or extrahepatic new lesions. The overall response status according to possible combinations of response classes in target and nontarget lesions with or without appearance of new lesions is shown in Table 1.

Radiologic responses were interpreted by both assessments protocols (EASL and mRECIST guideline) for the same target lesions on the same scan at the same time. To minimize possibility of false categorizations, all measurements were conducted by 2 independent observers (K.A. Kim and M.-J. Kim with 10 and 30 years of experience, respectively), both blinded to clinical data. Then, ultimately, final classifications made by consensus between 2 observers were adopted for analysis. Detailed data on agreement are provided in the Supplementary Table S1.

Statistical analysis

First, we examined intraindividual agreement rates, which was defined as the percentage of patients with the same result between response status, when considering only the longest 1, 2, 3, 4, or 5 lesions compared with all measurable lesions among whole population. Concordance between response assessments based on all measurable lesions in a given patient versus the longest 1, 2, 3, 4, or 5 lesions in the same patient was estimated using κ values. The strength of concordance based on κ values was interpreted as follows: $\kappa < 0.21$, poor; κ 0.21–0.40, fair; κ 0.41–0.60, moderate; κ 0.61–0.80, good; and $\kappa > 0.80$, excellent (16).

Next, we investigated prognostic values of treatment responses about OS, the primary endpoints of this study, when considering only the longest 1, 2, 3, 4, 5, and all target lesions, using EASL and mRECIST guidelines, respectively. Each prognostic value for OS was expressed as C-index, which was a natural extension of area under the receiver operating characteristic curve to analyze survival outcomes observed during longitudinal follow-up as a means of assessing discrimination with competing models (17).

OS was calculated as time interval between date of initiation of TACE and date of death or last follow-up, whereas progression-free survival (PFS) was calculated as time interval between date of initiation of TACE and date of progression or death. Survival time was estimated by Kaplan–Meier method, and survival difference between groups was assessed by log-rank test. For patients who underwent resection or liver transplantation after TACE, survival was censored at the time of surgery. Cox proportional HRs were calculated to test associations of clinical parameters with survivals.

Statistical analysis was conducted using SAS software version 9.1.3 (SAS Institute), and a 2-sided P value of less than 0.05 was considered statistically significant.

Table 1. Overall responses determined by evaluation of target, nontarget, and new lesions

Target lesions	Nontarget lesions	New lesions	Overall response
CR	CR	Absent	CR
CR	Non-CR or nonprogressive disease	Absent	PR
CR	Unmeasurable	Absent	PR
PR	Nonprogressive disease or unmeasurable	Absent	PR
Stable disease	Nonprogressive disease or unmeasurable	Absent	Stable disease
Progressive disease	Any response	Present or absent	Progressive disease
Any response	Progressive disease	Present or absent	Progressive disease
Any response	Any response	Present	Progressive disease

Table 2. Patients' baseline characteristics (*n* = 254)

Variables	Values
Age, y	60 (34–75)
Male gender	199 (78.3%)
Etiology	
HBV	161 (63.4%)
HCV	40 (15.7%)
Alcohol	30 (11.8%)
Others	23 (9.1%)
ECOG performance status	
0/1	163 (64.2%)/91 (35.8%)
Tumor number	
2	128 (50.4%)
3	53 (20.9%)
4	40 (15.7%)
$\geq$5	33 (13.0%)
Tumor size, cm	3.1 (1.5–9.7)
MELD score	4.98 (1–15.55)
AFP, ng/mL	78.0 (0.88–107,900)

NOTE: Values are expressed as median (range) or no. (%), unless indicated otherwise.

Abbreviations: HBV, hepatitis B virus; HCV, hepatitis C virus.

Results

Baseline characteristics

Baseline demographic and clinical characteristics are shown in Table 2. The median age of population (199 men and 55 women) was 60 (range, 34–75) years. All had Eastern Cooperative Oncology Group (ECOG) performance status of 0 or 1 and preserved liver function of Child–Pugh class A. The median diameter of the largest measurable lesion was 3.1 (range, 1.5–9.7) cm. The number of baseline measurable lesions was 2 in 128 patients (50.4%), 3 in 53 patients (20.9%), 4 in 40 (15.7%), and 5 or more in 33 (13.0%) patients. The median AFP level was 78.0 (range, 0.88–107,900) ng/mL.

Intraindividual agreement rates

Treatment responses according to EASL criteria are described in Table 3 (cells with agreement are in gray). In per-patient analysis, estimating a maximum number of 2, 3, 4, or 5 target lesions chosen in order of size resulted in an agreement rate of 95.3%, 98.4%, 100%, or 100% for target response and 94.9%, 98.4%, 100%, or 100% for overall response, respectively, compared with response estimating all measurable target lesions (Supplementary Fig. S2A and S2B). These results were significantly higher compared with agreement rates of 80.7% and 81.1% for target and overall response, respectively, obtained when considering only the

Table 3. Detailed responses according to maximum number of target lesions using EASL and mRECIST guideline

		EASL criteria								mRECIST							
		Target responses				Overall responses				Target responses				Overall responses			
		All targets				All targets				All targets				All targets			
Maximum number of targets		CR	PR	Stable disease	Progressive disease	CR	PR	Stable disease	Progressive disease	CR	PR	Stable disease	Progressive disease	CR	PR	Stable disease	Progressive disease
Up to 1	CR	93	31	1	0	90	31	1	0	93	31	1	0	90	31	1	0
	PR	0	59	6	0	0	59	6	0	0	48	8	1	0	48	8	1
	Stable disease	0	9	50	0	0	9	46	0	0	15	55	0	0	15	50	0
	Progressive disease	0	0	2	3	0	0	1	11	0	0	2	0	0	0	2	8
Up to 2	CR	93	4	0	0	90	5	0	0	93	4	0	0	90	4	0	0
	PR	0	91	2	0	0	90	2	0	0	83	3	0	0	83	3	0
	Stable disease	0	4	56	1	0	4	51	1	0	7	62	1	0	7	57	1
	Progressive disease	0	0	1	2	0	0	1	10	0	0	1	0	0	0	1	8
Up to 3	CR	93	2	0	0	90	2	0	0	93	2	0	0	90	2	0	0
	PR	0	96	0	0	0	96	0	0	0	89	1	0	0	89	1	0
	Stable disease	0	1	59	1	0	1	54	1	0	3	65	1	0	3	60	1
	Progressive disease	0	0	0	2	0	0	0	10	0	0	0	0	0	0	0	8
Up to 4	CR	93	0	0	0	90	0	0	0	93	0	0	0	90	0	0	0
	PR	0	99	0	0	0	99	0	0	0	94	0	0	0	94	0	0
	Stable disease	0	0	59	0	0	0	54	0	0	0	66	0	0	0	61	0
	Progressive disease	0	0	0	3	0	0	0	11	0	0	0	1	0	0	0	9
Up to 5	CR	93	0	0	0	90	0	0	0	93	0	0	0	90	0	0	0
	PR	0	99	0	0	0	99	0	0	0	94	0	0	0	94	0	0
	Stable disease	0	0	59	0	0	0	54	0	0	0	66	0	0	0	61	0
	Progressive disease	0	0	0	3	0	0	0	11	0	0	0	1	0	0	0	9

1 largest lesion with reference to response estimating all measurable targets (Supplementary Fig. S2A and S2B).

Treatment responses according to mRECIST are described in Table 3 (cells with agreement are in gray). Similarly, estimating a maximum number of 2, 3, 4, or 5 largest target lesions resulted in an agreement rate of 93.7%, 97.2%, 100%, or 100% for target response and 93.7%, 97.2%, 100%, or 100% for overall response, respectively, compared with response estimating all baseline measurable lesions (Supplementary Fig. S2C and S2D). These results were significantly higher compared with agreement rates of 77.2% and 77.2% for target and overall response, respectively, obtained when considering only the 1 largest lesion with reference to response estimating all measurable targets (Supplementary Fig. S2C and S2D).

Concordance between response evaluation methods

Kappa statistics showed "excellent" concordances between responses assessed using all measurable targets and using 2, 3, 4, or 5 largest target lesions by EASL criteria, as reflected by high κ values of 0.929 [95% confidence interval (CI), 0.889–0.968], 0.976 (95% CI, 0.953–0.999), 1.000 (95% CI, 1.000–1.000), or 1.000 (95% CI, 1.000–1.000) for target response and 0.924 (95% CI, 0.885–0.964), 0.977 (95% CI, 0.954–0.999), 1.000 (95% CI, 1.000–1.000), or 1.000 (95% CI, 1.000–1.000) for overall response, respectively. However, κ values between responses assessed using only the 1 longest target lesion and those using all measurable targets were 0.710 (95% CI, 0.639–0.781) and 0.723 (95% CI, 0.653–0.793) for target and overall response, respectively, showing only "good" levels of concordance.

Similarly, κ values between responses assessed using all measurable targets and using 2, 3, 4, or 5 largest target lesions by mRECIST were 0.905 (95% CI, 0.860–0.950),

0.958 (95% CI, 0.928–0.989), 1.000 (95% CI, 1.000–1.000), or 1.000 (95% CI, 1.000–1.000) for target response and 0.907 (95% CI, 0.864–0.951), 0.959 (95% CI, 0.930–0.989), 1.000 (95% CI, 1.000–1.000), or 1.000 (95% CI, 1.000–1.000) for overall response. Those between responses assessed using only the 1 longest target lesion and using all measurable targets were 0.657 (95% CI, 0.582–0.731) and 0.666 (95% CI, 0.593–0.739) for target and overall response, respectively, showing only "good" levels of concordance.

Prognostic value of mRECIST for predicting OS compared with EASL criteria

We calculated C-index to show prognostic values of each radiologic parameter in predicting OS. C-index by EASL criteria when measured using the longest 1, 2, 3, 4, 5, and all target lesions was similar in both target responses, ranging from 0.716 to 0.724 and overall responses ranged from 0.739 to 0.749 (Table 4).

Likewise, C-index by mRECIST when measured using the longest 1, 2, 3, 4, 5, and all target lesions was also similar in both target responses, ranging from 0.724 to 0.733 and overall responses ranged from 0.750 to 0.759 (Table 4).

Influence of radiologic response on OS

For target responses, using EASL and mRECIST guidelines, responders (subjects with CR and PR) had a significantly longer median OS than nonresponders (subjects with stable disease and progressive disease), regardless of maximum number of target lesions (Table 5). Likewise, for overall responses, using both criteria, similar results were obtained (Table 5). Figure 1 showed Kaplan–Meier analysis of OS using EASL criteria (A) and mRECIST (B).

Detailed HRs for both criteria are depicted in the Supplementary Table S2.

Table 4. Prognostic value for OS and PFS of EASL and mRECIST guidelines

Maximum number of targets	C-index for OS		C-index for PFS	
	Target responses by EASL criteria	Overall responses by EASL criteria	Target responses by EASL criteria	Overall responses by EASL criteria
Up to 1	0.716	0.739	0.711	0.710
Up to 2	0.718	0.744	0.709	0.723
Up to 3	0.719	0.744	0.707	0.724
Up to 4	0.724	0.749	0.711	0.728
Up to 5	0.724	0.749	0.711	0.728
All targets	0.724	0.749	0.711	0.728
	Target responses by mRECIST	Overall responses by mRECIST	Target responses by mRECIST	Overall responses by mRECIST
Up to 1	0.726	0.750	0.702	0.718
Up to 2	0.733	0.759	0.717	0.724
Up to 3	0.729	0.755	0.719	0.726
Up to 4	0.724	0.750	0.712	0.729
Up to 5	0.724	0.750	0.712	0.729
All targets	0.724	0.750	0.712	0.729

Table 5. Detailed median OS and adjusted HR from multivariate analysis

Maximum number of targets	Target response by EASL criteria		Overall responses by EASL criteria	
	OS, mo (responders vs. nonresponders)	Adjusted HRs (95% CI)	OS, mo (responders vs. nonresponders)	Adjusted HRs (95% CI)
Up to 1	40.5 vs. 20.7	2.079 (1.416–3.054)	40.8 vs. 20.7	2.270 (1.554–3.314)
Up to 2	40.5 vs. 21.2	1.765 (1.195–2.607)	40.8 vs. 21.2	1.937 (1.320–2.844)
Up to 3	40.5 vs. 21.2	1.733 (1.171–2.565)	40.8 vs. 20.7	1.904 (1.295–2.798)
Up to 4	40.5 vs. 21.2	1.733 (1.171–2.565)	40.8 vs. 21.2	1.904 (1.295–2.798)
Up to 5	40.5 vs. 21.2	1.733 (1.171–2.565)	40.8 vs. 21.2	1.904 (1.295–2.798)
All targets	40.5 vs. 21.2	1.733 (1.171–2.565)	40.8 vs. 21.2	1.904 (1.295–2.798)

	Target responses by mRECIST		Overall responses by mRECIST	
	OS, mo (responders vs. nonresponders)	Adjusted HRs (95% CI)	OS, mo (responders vs. nonresponders)	Adjusted HRs (95% CI)
Up to 1	40.8 vs. 25.3	2.287 (1.571–3.329)	40.8 vs. 22.9	2.491 (1.717–3.614)
Up to 2	40.8 vs. 21.3	2.108 (1.444–3.077)	40.8 vs. 21.2	2.303 (1.584–3.348)
Up to 3	40.8 vs. 21.3	2.243 (1.538–3.271)	40.8 vs. 21.2	2.447 (1.685–3.555)
Up to 4	40.8 vs. 23.3	2.163 (1.480–3.162)	40.8 vs. 21.3	2.363 (1.624–3.439)
Up to 5	40.8 vs. 23.3	2.163 (1.480–3.162)	40.8 vs. 21.3	2.363 (1.624–3.439)
All targets	40.8 vs. 23.3	2.163 (1.480–3.162)	40.8 vs. 21.3	2.363 (1.624–3.439)

NOTE: Adjusted HRs were calculated, adjusting other 3 clinical variables, tumor marker, tumor size, and tumor number from multivariate analysis.

Other independent factors influencing OS

Clinical variables other than radiologic responses, including age, sex, performance status, disease etiology, AFP level, model for end-stage liver disease (MELD) score, tumor number, and tumor size (largest lesion diameter) were analyzed in univariate analysis for OS (Supplementary Table S2). Among them, AFP level ($P < 0.001$), tumor number ($P < 0.001$), and tumor size ($P = 0.047$) were significant predictors of OS. Thus, these 3 variables were entered into a subsequent multivariate analysis along with one of the following radiologic parameters: target and overall responses by EASL criteria (responders vs. nonre-

sponders) and target and overall responses by mRECIST (responders vs. nonresponders) when assessed by 1, 2, 3, 4, 5, and all target lesions, respectively. Each multivariate analysis showed independent significance of radiologic response from each calculation method (all $P < 0.001$; Table 5). Detailed adjusted HRs from multivariate analysis are described in Table 5. Among remaining 3 clinical variables (AFP level, tumor size, and tumor number), AFP level and tumor number remained independently significant (both $P < 0.05$ in each multivariate analysis) along with radiologic response from each calculation method.
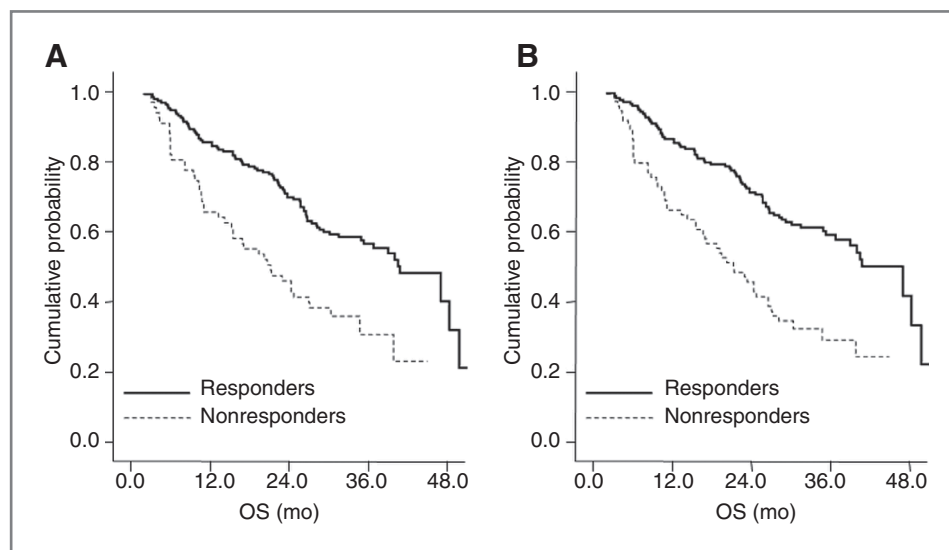

Figure 1. Kaplan–Meier analysis of OS using EASL criteria (A) and mRECIST (B). Representatively, overall treatment responses considering up to 2 target lesions were depicted, showing that responders had a significantly longer median OS compared with nonresponders (both $P < 0.001$).

Radiologic responses and PFS

We also calculated C-index to show prognostic values of each radiologic parameter in predicting PFS. Similar results were obtained; C-index by EASL criteria and mRECIST when measured using the longest 1, 2, 3, 4, 5, and all targets were similar in both target and overall responses (Table 4).

Furthermore, responders had the independently better PFS as compared with nonresponders; 10.1 (95% CI, 7.4–12.5) versus 6.4 (95% CI, 5.3–7.6) months from mRECIST and 10.1 (95% CI, 8.1–12.1) versus 6.6 (95% CI, 5.2–8.0) months from EASL criteria (both $P < 0.001$), when overall treatment responses considering up to 2 target lesions were evaluated. Similar results were obtained for all radiologic parameters; target and overall responses by EASL criteria (responders vs. nonresponders) and target and overall responses by mRECIST (responders vs. nonresponders) when assessed by 1, 2, 3, 4, 5, and all targets, respectively.

Discussion

To date, although response evaluation criteria, such as EASL and mRECIST, which take into consideration tumoral devascularization and/or necrosis using arterial enhancement of dynamic imaging get more popularity, investigations on optimum number of target lesions that should be evaluated for appropriate representation of overall tumor burden at baseline and subsequent follow-up are scarce. In response analyses, use of specific numbers of measurable targets should meet 2 major prerequisites: first, avoidance of interobserver variability and measurement errors, and second, avoidance of overburdening researchers with work in real clinical practice (18, 19). Clinical outcomes based on unoptimized target numbers can lead to misleading interpretations and erroneous clinical applications of a given treatment modality. Hence, we aimed to provide strong evidence to support the choice of the optimal number of target lesions that should be used in response evaluation following TACE for HCC based on analysis of OS, the most robust and unequivocal clinical parameter as a primary endpoint.

In our study, intraindividual agreement rates in response evaluation were similar when they were estimated using 2, 3, 4, or 5 targets chosen in order of size with reference to all HCC lesions, ranging from about 94% to 100% for both target and overall responses using EASL and mRECIST guidelines. However, agreement rates using only the 1 largest lesion were significantly low, ranging from approximately 77% to 81%. In a similar context, concordances between responses assessed using all measurable targets and using 2, 3, 4, or 5 largest targets by EASL and mRECIST guidelines were of "excellent" levels for both target and overall responses, whereas estimation of only the longest 1 target lesion showed only a "good" level of concordance with reference to responses using all measurable target lesions. These findings were consistent with results by Shim and colleagues (20), who indicated that evaluating the largest 2 lesions is generally the most useful procedure for assessing TACE responses under EASL and mRECIST guide-

lines. However, beyond above cross-sectional data, we took a step forward. We showed that prognostic abilities for predicting OS through a longitudinal study design, which was expressed as C-index, were practically equivalent regardless of maximum number of target lesions. Taken together, treatment responses might be assessed best using "mRECIST amendment in combination with 1.1 RECIST model" rather than "mRECIST in combination with original RECIST model (1.0 version)," especially in terms of convenience without compromising prognostic ability. In addition, radiologic response from all evaluation methods, including approach of using only 1 index lesion, was identified as the independent predictor for OS (all $P <$ 0.001). This phenomenon means that the largest 1 index lesion can reflect overall tumor burden as a representative to a substantially accurate level, even in cases of multiple tumors. Thus, only the 1 largest index lesion might be sufficient to provide prognostic information. That is, higher or lower concordance rates at cross-sectional level between responses assessing a specific number of target lesions and reference of assessing all target lesions do not necessarily mean higher or lower prognostic significances in predicting final survival outcomes. However, taken together, we cautiously propose that evaluating at least the largest 2 target lesions rather than only 1 should be recommended from the comprehensive standpoint that response evaluation at a cross-sectional level has been still an important endpoint in many clinical trials.

To our best knowledge, we first assessed prognostic values of each radiologic parameter through an optimized statistical method, C-index, in predicting survival outcomes according to number of target lesions. To date, most investigations have dealt with intermethod concordance between response assessment using specific number of target lesions and using all target lesions at a cross-sectional level (20). Second, we recruited a large number of patients undergoing TACE. Because TACE is one of the most commonly used locoregional treatments, our study may provide more standardized, generalized results applicable to such patients (21). In contrast, several key studies (8–10) included subjects who were heterogeneously treated with various treatment modalities, which may have biased interpretations of therapeutic responses and estimates of survival outcomes. Third, to minimize false categorization in evaluating responses, 2 independent radiologists interpreted radiologic responses and final classifications made by consensus between them were ultimately adopted. In our study, the high levels of concordance between the 2 readers were observed compared with other literatures (22–24). This is most likely because radiologists assessed responses using the given number of target lesions (i.e., up to 1 target, up to 2 targets, and so on) according to the main concept of this study. In this circumstance, interobserver variations that can be caused by considering the remaining lesions besides designated targets might have been prevented substantially. And, when anyone of the 2 radiologists requested further imaging modality for equivocal lesion, additional MRI scans were allowed, and these data were opened to both

radiologists. Along with providing radiologists, the detailed information about number of target lesions to be considered for response evaluation, this process might lessen obscurity and enhance concordance rates. Finally, we excluded patients with a previous history of HCC treatment. Considering that those with insufficient responses to previous treatments or recurrent disease might have had worse clinical outcomes, inclusion of only patients with treatment-naïve HCC might eliminate potential bias. In addition, we analyzed the best responses achieved in subsequent TACE sessions. Among 151 patients classified as PR or stable disease (using overall response by mRECIST) after the first TACE, about one thirds (42 patients) additionally achieved CR by means of so-called, "on-demand" protocols, and obviously, such patients had the better OS than remaining patients [47 (95% CI, 34.7–59.3) vs. 20.7 (95% CI, 14.5–26.9) months; $P < 0.001$, respectively]. According to best responses through "on demand" protocols among whole population, responders had the better OS compared with those with nonresponders [40.8 (95% CI, 35.8–45.8) vs. 13.3 (95% CI, 7.8–18.8) months; $P < 0.001$, respectively]. Taken together, based on our data, we validated rationales of mRECIST guidelines, which adopted recommendation for maximum number of target indicators suggested by RECIST (14) and suggest that EASL criteria can provide same efficacy without compromising prognostic ability even when using only 2 target lesions instead of all target lesions.

Consistent with our results, Riaz and colleagues (9) also showed that target response assessing only 1 index lesion is adequate for prediction of OS. In this study, differences in prognostic value between target response and overall response when estimating the same number of target lesions were statistically negligible for both EASL and mRECIST guidelines. However, Gillmore and colleagues (3) reported contradictory results: that target responses based on only the changes in designated target lesions did not appropriately reflect survival outcomes compared with overall responses because target response does not consider progression outside treated areas, which is a clinically relevant point in terms of survival outcome. Such a discrepancy between results by Gillmore and colleagues (3) and ours was most likely because progressive disease was relatively infrequent in the early phase of treatment courses in our treatment-naïve population.

Thus study had several limitations. A potential weakness of our study is lack of pathologic confirmation. A recent study by Golfieri and colleagues (25) suggested that tumors with complete lipiodol uptake are not always translated into "pathologic CR." However, as pathologic evaluations of treated lesions are available only in selected cases following resection, transplantation, or autopsy, one can suppose that response evaluation using dynamic MRI scan would be desirable for finer prognostification. Nevertheless, it is evident that complete lipiodol uptake remains a substantially favorable predictor, supported by many reports (12, 20, 26–29). Furthermore, in Republic of Korea, according to the reimbursement guidelines of the National Health Insurance Corporation, use of dynamic MRI scanning for response evaluation after HCC treatment remains limited to equivocal cases on dynamic CT scans. Second, this study focused only on a specific population where TACE was conducted. Because this population may not accurately represent entire HCC population, it is subject to selection bias. However, as TACE is the most commonly used locoregional therapy for HCC (30), this study may provide a basis for future research using another population treated with locoregional interventions or molecular-targeted agents (30). Further work is needed to determine whether this can also be translated to other treatment modalities. Third, OS, the primary endpoint of this study, can be confounded by other issues (e.g., underlying liver condition, post-TACE therapy), and thus, we attempted to minimize such biases by recruiting treatment-naïve subjects with well-preserved liver function and further analyzed clinical data from the viewpoint of PFS, with similar results. Fourth, radiologic evaluations are subject to inherent interobserver and intraobserver variation. And thus, in the current analysis, to minimize possibility of false categorizations, 2 independent radiologists interpreted radiologic responses and final classifications made by consensus between them were ultimately adopted. However, radiologic criteria have not shown a significant correlation with long-term outcomes in some cases (4, 31). To resolve these potential limitations of radiologic criteria, further researches are required to develop another method based on both radiologic and biologic criteria for an accurate monitoring clinical courses (32). Finally, excellent agreement rates when assessing up to 2 targets might be reduced to less than 90% in a subgroup with multiple tumors (≥ 5 tumors), as seen in the Supplementary Table S3. Although agreement when considering up to 2 targets was still greater than 85%, physicians should exercise caution when assessing treatment responses by considering the 2 largest targets in patients with multiple tumors (≥ 5 tumors).

In conclusion, our analysis suggests that prognostic values for predicting OS were similar regardless of maximum number of target lesions. However, assessing 2 largest target lesions rather than only 1 index lesion could be recommended considering high concordance at a cross-sectional level.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: B.K. Kim, S.U. Kim, J.Y. Park, S.H. Ahn, K.-H. Han
Development of methodology: B.K. Kim, S.U. Kim, S.H. Ahn
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): B.K. Kim, S.U. Kim, M.-J. Kim, K.A. Kim, D.Y. Kim, S.H. Ahn, K.-H. Han
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): B.K. Kim, S.U. Kim, M.-J. Kim, K.A. Kim, D.Y. Kim, J.Y. Park
Writing, review, and/or revision of the manuscript: B.K. Kim, S.U. Kim, K.A. Kim, D.Y. Kim, J.Y. Park
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): B.K. Kim, S.U. Kim, M.-J. Kim, K.-H. Han
Study supervision: B.K. Kim, S.U. Kim, J.Y. Park, S.H. Ahn, C.Y. Chon

References

1. Miller AB, Hoogstraten B, Staquet M, Winkler A. Reporting results of cancer treatment. Cancer 1981;47:207–14.
2. Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, 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: 205–16.
3. Gillmore R, Stuart S, Kirkwood A, Hameeduddin A, Woodward N, Burroughs AK, et al. EASL and mRECIST responses are independent prognostic factors for survival in hepatocellular cancer patients treated with transarterial embolization. J Hepatol 2011;55:1309–16.
4. Llovet JM, Ricci S, Mazzaferro V, Hilgard P, Gane E, Blanc JF, et al. Sorafenib in advanced hepatocellular carcinoma. N Engl J Med 2008; 359:378–90.
5. Lencioni R, Llovet JM. Modified RECIST (mRECIST) assessment for hepatocellular carcinoma. Semin Liver Dis 2010;30:52–60.
6. Bruix J, Sherman M, Llovet JM, Beaugrand M, Lencioni R, Burroughs AK, et al. Clinical management of hepatocellular carcinoma. Conclusions of the Barcelona-2000 EASL conference. European Association for the Study of the Liver. J Hepatol 2001;35:421–30.
7. Llovet JM, Di Bisceglie AM, Bruix J, Kramer BS, Lencioni R, Zhu AX, et al. Design and endpoints of clinical trials in hepatocellular carcinoma. J Natl Cancer Inst 2008;100:698–711.
8. Riaz A, Memon K, Miller FH, Nikolaidis P, Kulik LM, Lewandowski RJ, et al. Role of the EASL, RECIST, and WHO response guidelines alone or in combination for hepatocellular carcinoma: radiologic-pathologic correlation. J Hepatol 2011;54:695–704.
9. Riaz A, Miller FH, Kulik LM, Nikolaidis P, Yaghmai V, Lewandowski RJ, et al. Imaging response in the primary index lesion and clinical outcomes following transarterial locoregional therapy for hepatocellular carcinoma. JAMA 2010;303:1062–9.
10. Memon K, Kulik L, Lewandowski RJ, Wang E, Riaz A, Ryu RK, et al. Radiographic response to locoregional therapy in hepatocellular carcinoma predicts patient survival times. Gastroenterology 2011;141: 526–35.
11. Edeline J, Boucher E, Rolland Y, Vauleon E, Pracht M, Perrin C, et al. Comparison of tumor response by response evaluation criteria in solid tumors (RECIST) and modified RECIST in patients treated with sorafenib for hepatocellular carcinoma. Cancer 2012;118:147–56.
12. Shim JH, Lee HC, Kim SO, Shin YM, Kim KM, Lim YS, et al. Which response criteria best help predict survival of patients with hepatocellular carcinoma following chemoembolization? A validation study of old and new models. Radiology 2012;262:708–18.
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;2:1040–6.
14. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer 2009;45:228–47.
15. Korean Liver Cancer Study Group and National Cancer Center. [Practice guidelines for management of hepatocellular carcinoma 2009]. Korean J Hepatol 2009;15:391–423.
16. Altman D. Practical statistics for medical research. London: Chapman & Hall; 1991.
17. Zhou X, Obuchowski N, McClish D. Statistical methods in diagnostic medicine. New York: John Wiley and Sons, Inc; 2002.
18. Thiesse P, Ollivier L, Di Stefano-Louineau D, Negrier S, Savary J, Pignard K, et al. Response rate accuracy in oncology trials: reasons for interobserver variability. Groupe Francais d'Immunotherapie of the Federation Nationale des Centres de Lutte Contre le Cancer. J Clin Oncol 1997;15:3507–14.
19. Hopper KD, Kasales CJ, Van Slyke MA, Schwartz TA, TenHave TR, Jozefiak JA. Analysis of interobserver and intraobserver variability in CT tumor measurements. AJR Am J Roentgenol 1996;167:851–4.
20. Shim JH, Lee HC, Won HJ, Shin YM, Kim KM, Lim YS, et al. Maximum number of target lesions required to measure responses to transarterial chemoembolization using the enhancement criteria in patients with intrahepatic hepatocellular carcinoma. J Hepatol 2012;56: 406–11.
21. Han KH, Kudo M, Ye SL, Choi JY, Poon RT, Seong J, et al. Asian consensus workshop report: expert consensus guideline for the management of intermediate and advanced hepatocellular carcinoma in Asia. Oncology 2011;81(Suppl 1):158–64.
22. Marin D, Di Martino M, Guerrisi A, De Filippis G, Rossi M, Ginanni Corradini S, et al. Hepatocellular carcinoma in patients with cirrhosis: qualitative comparison of gadobenate dimeglumine-enhanced MR imaging and multiphasic 64-section CT. Radiology 2009;251:85–95.
23. Hwang J, Kim SH, Lee MW, Lee JY. Small ($</=$ 2 cm) hepatocellular carcinoma in patients with chronic liver disease: comparison of gadoxetic acid-enhanced 3.0 T MRI and multiphasic 64-multirow detector CT. Br J Radiol 2012;85:e314–22.
24. Kim SH, Lee J, Kim MJ, Jeon YH, Park Y, Choi D, et al. Gadoxetic acid-enhanced MRI versus triple-phase MDCT for the preoperative detection of hepatocellular carcinoma. AJR Am J Roentgenol 2009;192: 1675–81.
25. Golfieri R, Cappelli A, Cucchetti A, Piscaglia F, Carpenzano M, Peri E, et al. Efficacy of selective transarterial chemoembolization in inducing tumor necrosis in small (<5 cm) hepatocellular carcinomas. Hepatology 2011;53:1580–9.
26. Kim DY, Ryu HJ, Choi JY, Park JY, Lee DY, Kim BK, et al. Radiological response predicts survival following transarterial chemoembolisation in patients with unresectable hepatocellular carcinoma. Aliment Pharmacol Ther 2012;35:1343–50.
27. Riaz A, Kulik L, Lewandowski RJ, Ryu RK, Giakoumis Spear G, Mulcahy MF, et al. Radiologic-pathologic correlation of hepatocellular carcinoma treated with internal radiation using yttrium-90 microspheres. Hepatology 2009;49:1185–93.
28. Riaz A, Lewandowski RJ, Kulik L, Ryu RK, Mulcahy MF, Baker T, et al. Radiologic-pathologic correlation of hepatocellular carcinoma treated with chemoembolization. Cardiovasc Intervent Radiol 2010;33: 1143–52.
29. Maddala YK, Stadheim L, Andrews JC, Burgart LJ, Rosen CB, Kremers WK, et al. Drop-out rates of patients with hepatocellular cancer listed for liver transplantation: outcome with chemoembolization. Liver Transpl 2004;10:449–55.
30. Kim BK, Kim SU, Park JY, Kim do Y, Ahn SH, Park MS, et al. Applicability of BCLC stage for prognostic stratification in comparison with other staging systems: single centre experience from long-term clinical outcomes of 1717 treatment-naive patients with hepatocellular carcinoma. Liver Int 2012;32:1120–7.
31. Cheng AL, Kang YK, Chen Z, Tsao CJ, Qin S, Kim JS, et al. Efficacy and safety of sorafenib in patients in the Asia-Pacific region with advanced hepatocellular carcinoma: a phase III randomised, double-blind, placebo-controlled trial. Lancet Oncol 2009;10:25–34.
32. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: evolving considerations for PET response criteria in solid tumors. J Nucl Med 2009;50(Suppl 1):122S–50S.