

Interobserver variability in gross tumor volume delineation for hepatocellular carcinoma

Results of Korean Radiation Oncology Group 1207 study

Young Suk Kim¹ · Jun Won Kim² · Won Sup Yoon³ · Min Kyu Kang⁴ · Ik Jae Lee² · Tae Hyun Kim⁵ · Jin Hee Kim⁶ · Hyung-Sik Lee⁷ · Hee Chul Park⁸ · Hong Seok Jang⁹ · Chul Seung Kay⁹ · Sang Min Yoon¹⁰ · Mi-Sook Kim¹¹ · Jinsil Seong¹²

Received: 19 February 2016 / Accepted: 21 July 2016 / Published online: 18 August 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract

Purpose There has been increasing use of external beam radiotherapy for localized treatment of hepatocellular carcinoma (HCC) with both palliative and curative intent. Quality control of target delineation in primary HCC is essential to deliver adequate doses of radiation to the primary tumor while preserving adjacent healthy organs. We analyzed interobserver variability in gross tumor volume (GTV) delineation for HCC.

Patients and methods Twelve radiation oncologists specializing in liver malignancy participated in a multi-institutional contouring dummy-run study of nine HCC cases and independently delineated GTV on the same set of provided computed tomography images. Quantitative analysis was performed using an expectation maximization algorithm for simultaneous truth and performance level estimation (STAPLE) with kappa statistics calculating agreement

between physicians. To quantify the interobserver variability of GTV delineations, the ratio of the actual delineated volume to the estimated consensus volume (STAPLE), the ratio of the common and encompassing volumes, and the coefficient of variation were calculated.

Results The median kappa agreement level was 0.71 (range 0.28–0.86). The ratio of the actual delineated volume to the estimated consensus volume ranged from 0.19 to 1.93 (median 0.94) for all cases. The ratio of the common and encompassing volumes ranged from 0.001 to 0.56 (median 0.25). The coefficient of variation for GTV delineation ranged from 8 to 57 % (median 26 %).

Conclusion The interobserver variability in target delineation of HCC GTV in this study is noteworthy. Multi-institution studies involving radiotherapy for HCC require appropriate quality assurance programs for target delineation.

Young Suk Kim and Jun Won Kim contributed equally.

Electronic supplementary material The online version of this article (doi: [10.1007/s00066-016-1028-2](https://doi.org/10.1007/s00066-016-1028-2)) contains supplementary material, which is available to authorized users.

✉ Jinsil Seong
jsseong@yuhs.ac

¹ Department of Radiation Oncology, Jeju National University Hospital, Jeju National University School of Medicine, Jeju, Korea

² Department of Radiation Oncology, Gangnam Severance Hospital, Yonsei University College of Medicine, Seoul, Korea

³ Department of Radiation Oncology, Ansan Hospital, Korea University Medical Center, Ansan, Korea

⁴ Department of Radiation Oncology, Kyungpook National University School of Medicine, Daegu, Korea

⁵ Center for Liver Cancer, National Cancer Center, Goyang, Korea

⁶ Department of Radiation Oncology, Dongsan Medical Center, Keimyung University School of Medicine, Daegu, Korea

⁷ Department of Radiation Oncology, Dong-A University College of Medicine, Busan, Korea

⁸ Department of Radiation Oncology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea

⁹ Department of Radiation Oncology, The Catholic University of Korea College of Medicine, Seoul, Korea

¹⁰ Department of Radiation Oncology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea

¹¹ Department of Radiation Oncology, Korea Institute of Radiological & Medical Sciences, Seoul, Korea

¹² Department of Radiation Oncology, Severance Hospital, Yonsei University College of Medicine, 50-1 Yonsei-ro, 120-752 Seodaemun-gu, Seoul, Korea

Keywords Organs at risk · Radiotherapy · Chemoembolization · Radiofrequency ablation · Quality assurance

Interobservervariabilität in Abgrenzung zum makroskopischen Tumolvolumen für hepatozelluläre Karzinome

Ergebnisse der Korean Radiation Oncology Group 1207 Study

Zusammenfassung

Ziel Die externe kurative Strahlentherapie ist zunehmend bei der lokalisierten Behandlung hepatozellulärer Karzinome (HCC) in palliativer und kurativer Absicht in Gebrauch. Eine Qualitätskontrolle der Zielabgrenzung beim primären HCC ist entscheidend, um die passende Dosis für die Primärtumorbestrahlung zuzuführen und gesunde benachbarte Organe zu schonen. Wir analysierten die Interobservervariabilität in Abgrenzung zum makroskopischen Tumolvolumen (GTV) beim HCC.

Patienten und Methoden In einer institutsübergreifenden konturierenden Teststudie mit 9 HCC-Fällen beschrieben 12 Strahlentherapeuten das GTV anhand des gleichen Satzes von zur Verfügung gestellten Computertomogrammen. Die quantitative Analyse wurde unter Verwendung eines Expectation-Maximization-Algorithmus für die „Simultaneous Truth and Performance Level Estimation“ (STAPLE) in Verbindung mit einer zwischen den Ärzten vereinbarten Berechnung von Kappa-Statistiken durchgeführt. Um die Interobservervariabilität der GTV-Abgrenzung zu bestimmen, wurden das Verhältnis des tatsächlich abgegrenzten Volumens zum geschätzten Konsensvolumen (STAPLE), das Verhältnis des gemeinsamen Volumens zum umfassenden Volumen sowie der Variationskoeffizient berechnet.

Ergebnisse Das mediane Kappa-Agreement-Level betrug 0,71 (Spanne 0,28–0,86). Das Verhältnis des tatsächlich abgegrenzten Volumens zum geschätzten Konsensvolumen lag in sämtlichen Fällen zwischen 0,19 und 1,93 (Mittelwert 0,94). Das Verhältnis des gemeinsamen Volumens zum umfassenden Volumen lag zwischen 0,001 und 0,56 (Mittelwert 0,25). Der Variationskoeffizient für die GTV-Abgrenzung lag im Bereich von 8–57 % (Mittelwert 26 %).

Schlussfolgerung Die Interobservervariabilität in der Zielabgrenzung des HCC-GTV ist in dieser Studie bemerkenswert. Institutsübergreifende Studien für eine HCC-Strahlentherapie erfordern geeignete Qualitätssicherungsprogramme für die Zielabgrenzung.

Schlüsselwörter Risikoorgane · Strahlentherapie · Chemoembolisation · Radiofrequenzablation · Qualitätssicherung

There has been increasing use of external beam radiotherapy (EBRT) for localized treatment of hepatocellular carcinoma (HCC) with both palliative and curative intent [1]. EBRT has also been used in combination with transarterial chemoembolization (TACE) and radiofrequency ablation (RFA) [2]. Quality control of target delineation in primary HCC is essential to deliver adequate doses of radiation to the primary tumor, while preserving adjacent healthy organs [2]. However, there is no clear consensus on how to delineate target volumes in radiotherapy for HCC. In a recently published interobserver variability study, near-perfect or substantial agreement was seen in contouring gross tumor volume (GTV) in three case scenarios: HCC without portal vein thrombus (PVT), HCC with extensive PVT, and HCC with minor branch PVT (kappa values of 0.82, 0.80, and 0.71, respectively) [3]. It still remains to be verified whether contouring agreement would still hold when a variety of clinical cases are presented.

The purpose of this multi-institution dummy-run study was to quantify the interobserver variability in GTV delineation for nine HCC cases presenting with various tumor characteristics.

Patients and methods

In November 2012, the Korean Radiation Oncology Group (KROG) established a multicenter study to analyze the interobserver variability in GTV delineation for primary HCC in EBRT (KROG 1207). Twelve radiation oncologists specializing in HCC EBRT from 12 institutions participated in the study.

Diagnostic liver magnetic resonance imaging (MRI), multiphasic computed tomography (CT; Somatom Sensation 64; Siemens Medical Solutions, Forchheim, Germany), and planning CT images were provided through the study website using Digital Imaging and Communications in Medicine (DICOM) format files. All datasets were fully anonymized before being delivered to the participating centers.

This study protocol conformed to the ethical guidelines outlined in the 1975 Declaration of Helsinki and was approved by the Institutional Review Boards of participating institutions.

Case histories

Clinical information of nine selected patients is presented in supplementary table 1. Patients 1, 2, 3, and 6 had been treated with TACE prior to EBRT. Right main PVT was observed in patients 3, 4, 5, and 9. An infiltrative and ill-defined tumor margin was present in patients 4, 5, 6, and 9.

Diagnostic multiphase CT images were provided in four cases, whereas MRI images were provided for all patients.

Image acquisition

The planning CT scan was obtained using a standard acquisition protocol (arterial phase contrast-enhanced, supine position, both arms above the head, free breathing, and slice thickness of 5 mm) using a SOMATOM Sensation Open (Siemens Medical, Erlangen, Germany) CT scanner [4]. The slice thickness was 3 mm in patients 1 and 2. Two-dimensional (2D) fluoroscopy was acquired in every HCC patient to evaluate respiratory movement of the diaphragm. Four-dimensional (4D) CT simulation using Respiratory Gating System 3.0 (Anzai Medical, Tokyo, Japan) was acquired in the case of hypofractionated radiotherapy or respiratory movement of the diaphragm on 2D fluoroscopy over 1–1.5 cm. An abdominal compression device was used in the case of hypofractionated radiotherapy or respiratory movement of the diaphragm on 2D fluoroscopy over 1–1.5 cm. If there was no improvement of respiratory movement of the diaphragm on 2D fluoroscopy with an abdominal compression device, an abdominal compression device was not applied in that case. In the present study, four cases (cases 1, 2, 4, and 6) underwent 4D-CT simulation. An abdominal compression device was used in six cases (cases 1, 2, 3, 5, 6, and 9). Diagnostic liver MRI (MAGNETOM Tim Trio, Siemens Healthcare; Achieva, Philips Healthcare, Eindhoven, Netherlands) in gadolinium-enhanced T1-weighted volumetric interpolated breath-hold examination (VIBE) at arterial phase (20–35 s) and T1-weighted VIBE at hepatobiliary delayed phase (20 min) were provided for all nine patients.

Target delineation

Patient histories and the official radiographic interpretation of MRI were provided to the panelists to aid decision making in identification of GTV. All physicians obtained detailed instruction, which is provided in the KROG 1207 study protocol, for definition of the GTV as follows:

1. The visible extent of hepatic parenchyma primary tumor and portal vein invading lesion in the soft tissue window.
2. Fusion of MRI and planning CT was recommended.
3. Respiratory motion of the gross tumor was not considered in GTV delineation.

It was possible to contour the primary tumor and portal vein-invading lesion as separate structures. The limits for the soft tissue window levels could be chosen by the participants. The magnification factor was left to the physician's discretion. Sagittal and coronal CT reconstructions were

available to allow orientation in the craniocaudal direction [5]. Contouring was performed independently without the information of other physicians. Normal structures (liver, stomach, duodenum, bowel, spinal cord, and kidney) were provided to the panelists. Physicians also completed a short survey of their clinical experience with irradiation of HCC.

Interobserver variability analysis

The completed contours were centrally collected in DICOM format and loaded onto the Eclipse™ (Varian, Palo Alto, CA, USA) treatment planning system. Each case was then imported into the Computational Environment for Radiotherapy Research program (CERR version 4.4) run through MATLAB version 7.8 (MathWorks Inc., Natick, MA, USA) for analysis [6]. Quantitative analysis on expert agreement was performed using an expectation maximization algorithm for Simultaneous Truth and Performance Level Estimation (STAPLE) [7]. Kappa statistics were used to assess agreement between contouring observers [8]. Perfect agreement between physicians would equate to a kappa value of 1, and a value of 0 represents no agreement. Estimated consensus contours for the GTV were generated at a 95% confidence level (S95) based on STAPLE analysis from the 12 overlaid contours from each case. The contours created by the radiation oncologist (J.S.) with the highest volume of cases in conjunction with a diagnostic radiologist were defined as the reference contours. The S95 of each case was compared to the reference contours and scored using Dice's coefficient, defined as the intersection of volumes/average volume as a similarity metric (volumes in cm³), with higher values indicating greater agreement (median 0.83, range 0.41–0.93). The S95 was then finalized as the consensus contour [3]. Ratios of the true contoured GTV of individual observers to the S95 for each of case were calculated. The dimensionless coefficient of variation (%), defined as the ratio between the standard deviation and the mean, was calculated to measure the relative data scatter with respect to the mean [9]. The ratio between a common volume (i. e., a volume on which all observers agreed) and an encompassing volume (i. e., the smallest volume that included all observers' volumes) was also calculated [10].

Results

Four panelists adhered to the recommendation to use CT-MRI fusion for delineating all 12 cases. Among these panelists, two used the deformable registration method in registering the CT images of the liver with those from the MRI. The rest of the panelists did not use CT-MRI fusion; however, all panelists thoroughly reviewed the MRI images before delineating the GTV. The main reason for not us-

Table 1 Variations in the contouring of gross tumor volume (GTV) of each case

Statistics	1	2	3	4	5	6	7	8	9
V_{\max} (ml)	7	19	547	100	288	103	538	160	2365
V_{\min} (ml)	0.7	9	259	35	204	25	416	53	1274
V_{avg} (ml)	3	13	366	69	238	52	492	101	1753
V_{SD} (ml)	± 1.8	± 3	± 106	± 20	± 23	± 26	± 43	± 26	± 317
V_{\max}/V_{\min}	10.1	2.1	2	2.8	1.4	4	1.3	3	1.6
Comm (ml)	0.01	1	159	20	165	8	355	46	1207
Encom (ml)	9	29	624	158	358	126	632	177	2544
C/E	0.001	0.04	0.25	0.13	0.46	0.07	0.56	0.26	0.47
COV (%)	57	25	28	30	10	50	8	26	18
STAPLE volume (ml)	3	15	390	72	231	80	498	100	1668
Kappa	0.28	0.58	0.71	0.60	0.82	0.60	0.86	0.71	0.81

SD standard deviation, V_{\max} maximum volume, V_{\min} minimum volume, V_{\max}/V_{\min} ratio of V_{\max} to V_{\min} , V_{avg} average volume, *Comm* the largest volume common to all the individual GTVs, *Encom* the smallest volume encompassing all the individual GTVs, *C/E* ratio of common to encompassing volume, *COV* coefficient of variation (SD/mean), *STAPLE* simultaneous truth and performance level estimation

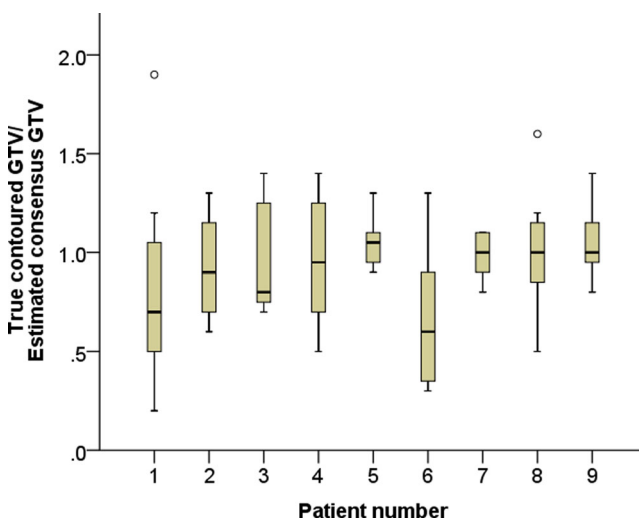


Fig. 1 Box-and-whisker plots of the ratio of true contoured gross tumor volume (GTV) to the 95 % confidence level (S95) for each of nine patients. The box spans the first to the third quartile; the *line* inside each *box* shows the median values and the *upper and lower whiskers* indicate the range

ing CT-MRI fusion for GTV delineation was the positional difference of the liver between CT and MRI images. One panelist contoured the primary lesion and vessel-invading lesion as separate structures. The coefficient of variation ranged from 8 to 57 % (median 26 %). The median kappa agreement level was 0.71 (range 0.28–0.86; Table 1). Ratios of the true contoured volume to the S95 for each patient ranged from 0.19 to 1.93 (median 0.94; Fig. 1).

The largest variability was observed in patient 1. The recurrent lesion was adjacent to the previous TACE site (lipiodol uptake site). Four observers contoured the GTV including the lipiodol uptake site, and the others contoured the GTV excluding the lipiodol uptake site (supplementary

figure 1). Diagnostic liver multiphase CT images were not provided. Due to the small volume, small inaccuracies in outlining the tumor had an amplified impact on the ultimately delineated GTV.

In patient 2, residual tumor after incomplete TACE did not show clear enhancement on the contrast-enhanced arterial phase planning CT scan. Coregistered MRI images did not align well with planning CT images due to differences in patient positioning during image acquisition, resulting in noteworthy liver deformation (supplementary figure 2).

In patient 4, ill-defined and infiltrative perivascular mass formation (in the right upper anterior PVT) likely contributed to the variability. One observer (blue line) delineated a triangular zone of subtle arterial enhancement (nearly same as normal liver parenchyma) that was narrow medially and widened laterally adjacent to PVT on MRI (Fig. 2).

In patient 6, two observers delineated the previous RFA site (low signal intensity on delayed hepatobiliary phase on MRI) adjacent to the recurrent tumor, medially and posteriorly. There was a site of lipiodol uptake from previous TACE adjacent to the recurrent tumor. Five observers delineated the GTV including this lipiodol uptake site. The tumor in patient 6 was located in segments 2 and 3, adjacent to the pyloric portion of the stomach inferiorly, and adjacent to the heart superiorly. Additional inaccuracies may have been attributed to partial volume averaging, which enlarges with increasing slice thickness (Fig. 3).

Variations in delineation of the distal aspect of a tumor thrombus were observed in patient 3 with extensive PVT (supplementary figure 3) and patient 9 with inferior vena cava invasion (supplementary figure 7). It was difficult to discriminate the extent of benign thrombi contiguous with malignant thrombi.

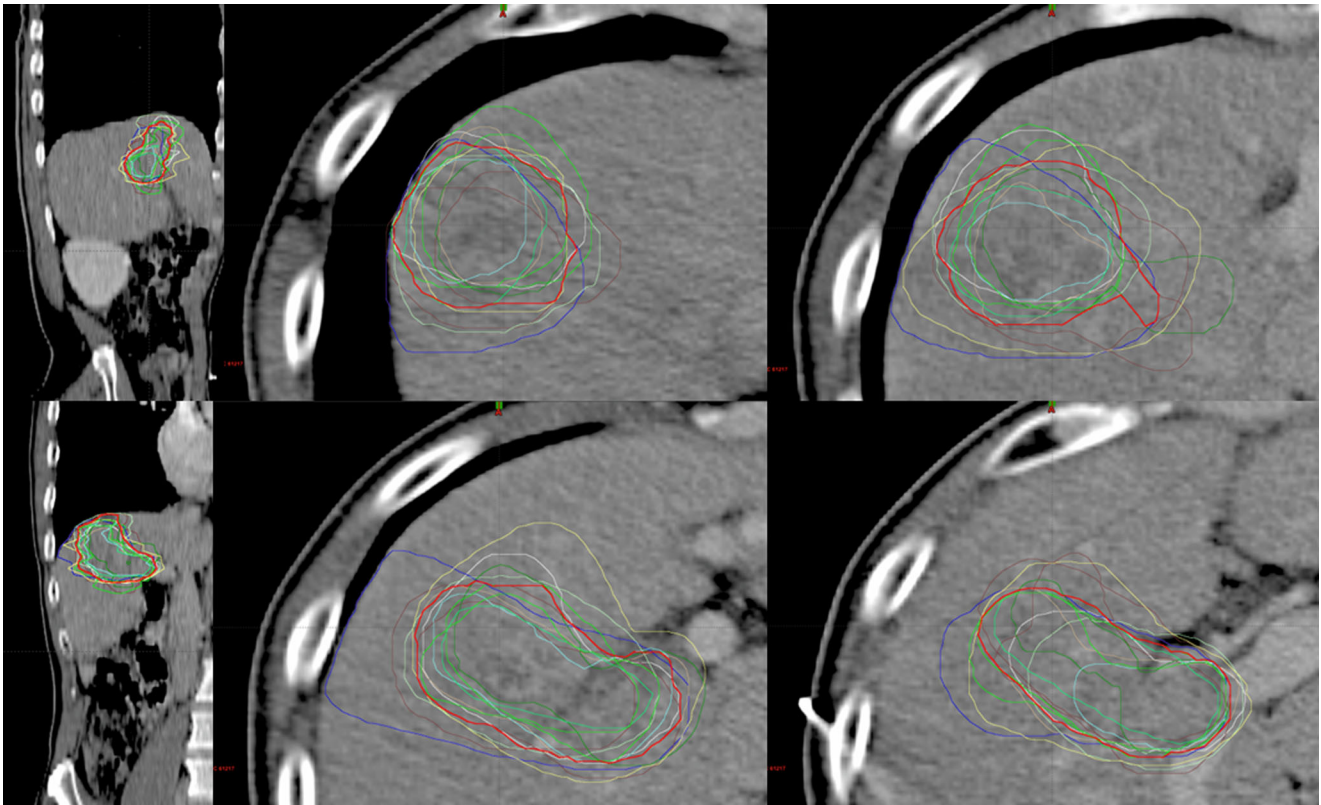


Fig. 2 Patient 4. The 95 % confidence level (S95) is shown in *thicker red line*

The smallest variability was observed for patient 7. The tumor had a relatively well-demarcated margin. Diagnostic multiphase CT images were provided (supplementary figure 5).

Observers 3, 5, 7, and 9 contoured the GTV 20 % smaller than the S95 in six, four, six, and four cases, respectively. Observer 3 contoured the GTV 10 % smaller than the S95 in all nine cases. Observers 1, 4, and 8 contoured the GTV 20 % larger than the S95 in five, four, and four cases, respectively. This tendency was irrespective of the extent of lifetime irradiation experience of HCC cases. Observers 1, 3, 7, 8, and 9 had experience with more than 100 cases of HCC irradiation in their lifetime.

Discussion

All institutions that participated in the current study are located in HCC-endemic areas, and the radiation oncologists had abundant clinical experience in treating HCC. Nonetheless, our results indicate substantial variation in GTV delineation among radiation oncologists from different institutions, especially in four cases.

Factors that may have been associated with the largest variation include absence of diagnostic multiphase CT, tumor location adjacent to the previous treatment with TACE

or RFA, ill-defined infiltrative tumor margin, extensive PVT, and small tumor size.

Diagnostic multiphase CT images were provided for patients 2, 5, 7, and 9. The coefficient of variation for GTV in these four patients ranged from 8 to 25 % (mean 15 %). The coefficient of variation for GTV ranged from 26 to 57 % (mean 38 %) in the other five cases.

Hong et al. reported near-perfect or substantial agreement in contouring the primary HCC GTV [3]. In contrast to this, our study included four patients with previous treatment with TACE and RFA. In patients 1 and 6, tumors were located adjacent to the previous treatment area with TACE or RFA. Previous treatment with TACE or RFA likely contributed to the greater variability in these patients.

Lieven et al. reported the influence of size and window setting on measured volume in a phantom study. In his study, the selection of window center had an important influence on the measured volume. The effect of window setting on measured object volume was much more important in the case of small object size. This may be explained by the predominant effect of volume averaging in small objects [11]. In the future, provision of specific contouring guidelines will reduce interobserver variability, such as requirement of diagnostic multiphase CT images and window setting. Target contouring education before multi-in-

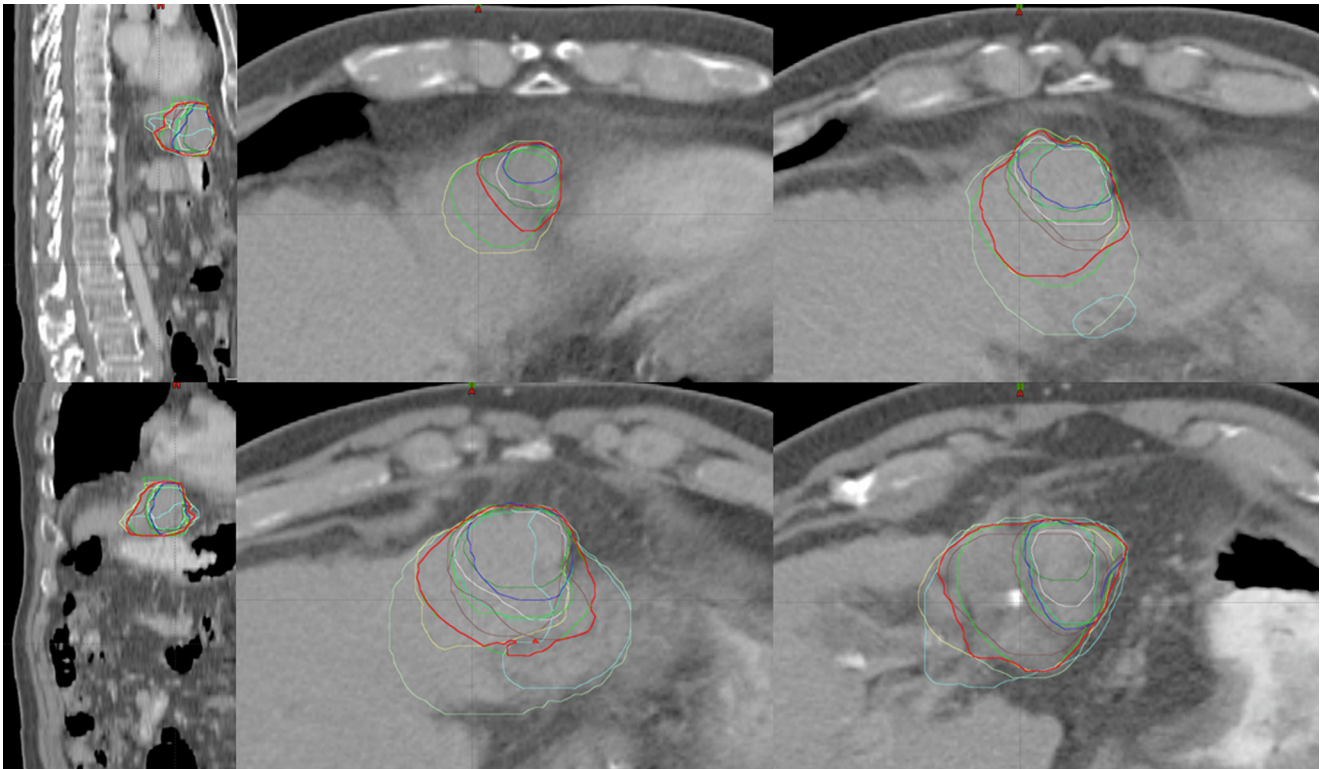


Fig. 3 Patient 6. The 95 % confidence level (S95) is shown in *thicker red line*. Lipiodol deposit after transarterial chemoembolization (TACE) presented with high-density nodule

stitution study and consensus meeting discussion may help in improving frequent errors and disagreements [12].

Another issue is utilization of imaging studies in target delineation. Coregistration of MRI improved the accuracy of GTV delineation greatly in radiotherapy for many tumors, including prostate cancer [13], head and neck cancer [14], and brain tumors [15]. MRI improved diagnosis and definition of HCC tumor extent [16], and MRI images were provided to assist target delineation for all patients in the current study. However, unlike delineating tumors in the brain, respiration-dependent motion of the liver poses limits to the use of MRI for HCC GTV delineation. Several motion management strategies are employed in order to compensate for the respiratory-dependent motion of the liver, including abdominal compression to reduce the breathing motion, treating during repeated breath holds, gating radiotherapy, and real-time tumor tracking [17]. In designing a radiotherapy protocol for a multi-institution study, it is important to minimize technical requirements, while maintaining quality control of the treatment protocol. Investigators need to be aware of the technologies available at participating institutions.

Recent studies have shown that the MRI deformable registration technique for liver can improve the accuracy of tumor delineation [18]. Further investigations of liver MRI deformable registration may help in improving the accuracy

of target delineation related to the respiratory motion and MRI acquisition position.

Consultation with a diagnostic hepatobiliary radiologist is encouraged to improve the accuracy of target delineation [3]. In particular, discussion with a diagnostic radiologist may help in targeting recurrent tumor adjacent to the previous TACE and RFA treatment area, as well as tumor with extensive PVT.

Introduction of 4D-CT has improved respiratory management of tumors in the abdomen and thorax [19, 20]. However, acquiring 4D-CT images with intravenous (IV) contrast is not practical, since 4D-CT may take up to 1 min or more and administration of multiple doses of IV contrast may be necessary. In a protocol developed by Beddar et al., CT scans for treatment planning were acquired free of contrast agents while the patient breathed freely. 4D-CT scanning was performed only in the region containing the liver in cine mode for at least one complete breathing cycle, while the IV contrast was synchronized with the 3D image acquisition [21]. However, this protocol is unlikely to be universally accepted since many institutions rely on conventional planning CT images.

The current study has several limitations. Diagnostic multiphase CT images were not provided for all sample patients. The impact of multiphase CT images may have been more pronounced for tumors with less clear margins

and may have further reduced interobserver variability. Detailed treatment histories were also not provided for the cases for practical reasons. However, many clinicians take into consideration treatment history and the current status of the patient and disease. The quality of planning CT images was not optimal. To reduce interobserver variability for future prospective studies, a protocol for acquiring high-quality planning CT images in addition to implementing the breath-hold technique and reduced slice thickness, is required. Lastly, we did not evaluate target delineation of the clinical target volume (CTV) of primary HCC. Delineation of CTV can vary among physicians. Future research is needed to investigate interobserver variability of CTV for primary HCC.

Hong et al. were able to provide consensus guidelines for workflow in the GTV definition of HCC [3]. We added the following recommendations based on the results of our study.

1. Small slice thickness (≤ 3 mm) is recommended for simulation CT, especially for small HCCs.
2. Use of a respiratory motion-restricting method, such as abdominal compression, is recommended in addition to 4D-CT for patients with large (over 1–1.5 cm) respiratory movement.
3. Diagnostic multiphase CT imaging is essential for contouring GTV.
4. MRI fusion is strongly recommended for HCC with features including hypovascularity, ill-defined tumor margin, and PVT. Vascular structures can be used as reference points for image fusion.
5. Consensus on inclusion of previous TACE site in GTV is inconclusive; however, inclusion of initial TACE site is recommended when marginal recurrence was observed. A multidisciplinary team approach is recommended for deciding the treatment volume.
6. Consensus meeting is encouraged to discuss guidelines for contouring GTV, CTV, and PTV prior to developing multi-institutional study protocols involving radiotherapy for HCC.

Conclusion

In the current study, the interobserver variability in the target delineation of GTV for primary HCC was noteworthy. To reduce this variability, the authors plan to evaluate the variability further and come up with a consensus guideline as an extension to this study in the near future. In designing a multi-institution study of radiotherapy for primary HCC, clear guidelines and appropriate quality assurance of target delineation are necessary.

Acknowledgements This work was supported by a grant (0620390) from the National R&D Program for Cancer Control, Ministry of Health and Welfare, Republic of Korea.

Compliance with ethical guidelines

Conflict of interest Y.S. Kim, J.W. Kim, W.S. Yoon, M.K. Kang, I.J. Lee, T.H. Kim, J.H. Kim, H.-S. Lee, H.C. Park, H.S. Jang, C.S. Kay, S.M. Yoon, M.-S. Kim, and J. Seong declare that they have no competing interests.

Ethical standards All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study.

References

1. Park J, Won, Chen M et al (2015) Global patterns of hepatocellular carcinoma management from diagnosis to death: the BRIDGE Study. *Liver Int* 35:2155–2166
2. Seong J (2009) Challenge and hope in radiotherapy of hepatocellular carcinoma. *Yonsei Med J* 50:601–612
3. Hong TS, Bosch WR, Krishnan S et al (2014) Interobserver variability in target definition for hepatocellular carcinoma with and without portal vein thrombus: radiation therapy oncology group consensus guidelines. *Int J Radiat Oncol Biol Phys* 89:804–813
4. Genovesi D, C afaro GA, Vinciguerra A et al (2011) Interobserver variability of clinical target volume delineation in supradiaphragmatic Hodgkin’s disease: a multi-institutional experience. *Strahlenther Onkol* 187:357–366
5. Vorwerk H, Beckmann G, Bremer M et al (2009) The delineation of target volumes for radiotherapy of lung cancer patients. *Radiother Oncol* 91:455–460
6. Cox BW, Spratt DE, Lovelock M et al (2012) International Spine Radiosurgery Consortium consensus guidelines for target volume definition in spinal stereotactic radiosurgery. *Int J Radiat Oncol Biol Phys* 83(e597):e605
7. Warfield SK, Zou KH, Wells WM (2004) Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans Med Imaging* 23:903–921
8. Sim J, Wright C (2005) The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther* 85:257–268
9. Villeirs GM, Van Vaerenbergh K, Vakaet L et al (2005) Interobserver delineation variation using CT versus combined CT + MRI in intensity-modulated radiotherapy for prostate cancer. *Strahlenther Onkol* 181:424–430
10. Weiss E, Hess CF (2003) The impact of gross tumor volume (GTV) and clinical target volume (CTV) definition on the total accuracy in radiotherapy theoretical aspects and practical experiences. *Strahlenther Onkol* 179:21–30
11. Van Hoe L, Haven F, Bellon E et al (1997) Factors influencing the accuracy of volume measurements in spiral CT: a phantom study. *J Comput Assist Tomogr* 21:332–338
12. Eriksen J, Salembier C, Rivera S et al (2014) Four years with FALCON – an ESTRO educational project: achievements and perspectives. *Radiother Oncol* 112:145–149

13. Rasch C, Barillot I, Remeijer P et al (1999) Definition of the prostate in CT and MRI: a multi-observer study. *Int J Radiat Oncol Biol Phys* 43:57–66
14. Rasch C, Keus R, Pameijer FA et al (1997) The potential impact of CT-MRI matching on tumor volume delineation in advanced head and neck cancer. *Int J Radiat Oncol Biol Phys* 39:841–848
15. Weltens C, Menten J, Feron M et al (2001) Interobserver variations in gross tumor volume delineation of brain tumors on computed tomography and impact of magnetic resonance imaging. *Radiother Oncol* 60:49–59
16. Voroney J, Brock K, Eccles C et al (2006) Prospective comparison of computed tomography and magnetic resonance imaging for liver cancer delineation using deformable image registration. *Int J Radiat Oncol Biol Phys* 66:780–791
17. Dawson LA, Guha C (2008) Hepatocellular carcinoma: radiation therapy. *Cancer J* 14:111–116
18. Xu H, Gong G, Wei H et al (2014) Feasibility and potential benefits of defining the internal gross tumor volume of hepatocellular carcinoma using contrast-enhanced 4D CT images obtained by deformable registration. *Radiat Oncol* 9:221
19. Pan T, Lee T, Rietzel E et al (2004) 4D-CT imaging of a volume influenced by respiratory motion on multi-slice CT. *Med Phys* 31:333–340
20. Hong TS, DeLaney TF, Mamon HJ et al (2014) A prospective feasibility study of respiratory-gated proton beam therapy for liver tumors. *Pract Radiat Oncol* 4:316–322
21. Beddar AS, Briere TM, Balter P et al (2008) 4D-CT imaging with synchronized intravenous contrast injection to improve delineation of liver tumors for treatment planning. *Radiother Oncol* 87:445–448