

Scientific Article

Interobserver Variability of Gross Tumor Volume Delineation for Colorectal Liver Metastases Using Computed Tomography and Magnetic Resonance Imaging



Cora Marshall, MSc,^{a,b,*} Pierre Thirion, MD,^{b,c} Alina Mihai, MD,^b
John G. Armstrong, MD,^b Seán Cournane, PhD,^a Darina Hickey, MSc,^b
Brendan McClean, PhD,^{a,c} and John Quinn, PhD^a

^aSchool of Physics, University College Dublin, Dublin, Ireland; ^bBeacon Hospital, Dublin, Ireland; ^cSt Luke's Radiation Oncology Network, Dublin, Ireland

Received October 27, 2021; accepted June 28, 2022

Abstract

Purpose: The purpose of this study was to evaluate the interobserver variability in the contouring of the gross tumor volume (GTV) on magnetic resonance (MR) imaging and computed tomography (CT) for colorectal liver metastases in the setting of SABR.

Methods and Materials: Three expert radiation oncologists contoured 10 GTV volumes on 3 MR imaging sequences and on the CT image data set. Three metrics were chosen to evaluate the interobserver variability: the conformity index, the DICE coefficient, and the maximum Hausdorff distance (HDmax). Statistical analysis of the results was performed using a 1-sided permutation test.

Results: For all 3 metrics, the MR liver acquisition volume acquisition (MR LAVA) showed the lowest interobserver variability. Analysis showed a significant difference ($P < .01$) in the mean DICE, an overlap metric, for MR LAVA (0.82) and CT (0.74). The HDmax that highlights boundary errors also showed a significant difference ($P = .04$) with MR LAVA having a lower mean HDmax (7.2 mm) compared with CT (5.7 mm). The mean HDmax for both MR single shot fast spin echo (SSFSE) (19.3 mm) and diffusion weighted image (9.5 mm) showed large interobserver variability with MR SSFSE having a mean HDmax of 19.3 mm. A volume comparison between MR LAVA and CT showed a significantly higher volume for small GTVs ($<5 \text{ cm}^3$) when using MR LAVA for contouring in comparison to CT.

Conclusions: This study reported the lowest interobserver variability for the MR LAVA, thus indicating the benefit of using MR to complement CT when contouring GTV for colorectal liver metastases.

© 2022 Published by Elsevier Inc. on behalf of American Society for Radiation Oncology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Stereotactic ablative radiation therapy (SABR) is an external beam radiation therapy technique that uses precise targeting to deliver high doses of radiation capable of ablating tumors directly.¹ Treating primary or secondary liver malignancies with these ablative doses has become possible

Sources of support: This work had no specific funding.

Disclosures: We have no conflicts of interest to disclose.

Data sharing statement: Research data are not available at this time.

*Corresponding author: Cora Marshall, MSc; E-mail:

marshallcora@gmail.com

<https://doi.org/10.1016/j.adro.2022.101020>

2452-1094/© 2022 Published by Elsevier Inc. on behalf of American Society for Radiation Oncology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

with the emergence of image guided radiation therapy and respiratory management. The delivery of radiation to reduced planning target volumes (PTVs) allows for functional liver, away from the target area, to be spared.²

As a result, SABR is increasingly used in the management of liver metastases, with clinical series reporting promising 2-year local control rates, of approximately 90%.³ Studies have shown that liver SABR could have a major role in treating colorectal cancer patients, for whom the liver is the dominant metastatic site. In some cases, particularly patients with oligometastatic disease^{4,5} when there are a limited number of tumors, up to 5 in the liver, the aim is to eradicate the disease completely in liver.

Due to the steep dose gradients in SABR treatments, the accurate determination of the gross tumor volume (GTV) is a crucial step. However, it is widely accepted that this step of delineation of the GTV by the radiation oncologist is subject to interobserver variability.⁶ While numerous studies have evaluated interobserver variability, a recent review of 119 studies⁷ has identified only one that has examined interobserver variability in liver cancer.

In liver SABR, the precise delineation of the GTV is challenging due to the poor soft tissue contrast of computed tomography (CT) and the limited literature identifying pathologic correlation with radiologic features. Despite these limitations, CT remains the clinical standard for volume delineation in radiation therapy; however, other modalities are increasingly being utilized and showing promise. Magnetic resonance (MR) imaging (MRI) is now considered the gold standard for delineation of brain tumors² for stereotactic treatments, offering superior soft-tissue contrast to that of CT imaging. Furthermore, the use of MRI for the delineation of abdominal tumors has also been reported to be increasing.⁶

According to International commission on Radiation Units and Measurements (ICRU) 83,⁸ a clinical margin is added to the GTV to determine the PTV. Random and systematic uncertainties do not have an equal effect on the dose distribution. Random errors cause a blurring of the dose distribution where systematic errors cause a shift of the cumulative dose distribution. Interobserver variability is considered a systematic error. The reduction in such errors should be optimized to prevent inadvertent irradiation of normal tissues, particularly in high-dose treatments.

The primary objective of this study was to evaluate the interobserver delineation variation for colorectal liver metastases for SABR when using CT-based GTV delineation and MR-based delineation for a number of MR sequences. In addition, we aimed to establish which MR sequence yielded the lowest interobserver variability.

Methods and Materials

The study was approved by the institutional clinical audit committee of the institution.

Table 1 Information on the GTVs delineated, the segment of the liver, the estimated size of the tumor by the radiologist, the timing of the image after contrast injection, whether a DWI was available, and if a contrast-enhanced CT was possible

| GTV | Liver segment | Size (cm) | MR LAVA contrast timing (s) | MR DWI |
|--------|---------------|-----------|-----------------------------|--------|
| GTV 1 | 2 | 1.3 | 70 | Yes |
| GTV 2 | 7 | 2 | 130 | Yes |
| GTV 3 | 6 | 4 | 70 | No |
| GTV 4 | 5 | 3.5 | 70 | No |
| GTV 5 | 6 | 2 | 70 | No |
| GTV 6 | 8 | 4.5 | 130 | No |
| GTV 7 | 6 | 1.4 | 70 | Yes |
| GTV 8 | 7 | 2 | 70 | Yes |
| GTV 9 | 7 | 2 | 70 | Yes |
| GTV 10 | 7 | 2 | 70 | Yes |

Abbreviations: CT = computed tomography; DWI = diffusion weighted image; GTV = gross tumor volume; LAVA = liver acquisition volume acquisition; MR = magnetic resonance.

Patient database and eligibility

An anonymized database was created from 7 patients with metastatic colorectal cancer having attended our institution for liver SABR, representing a total of 10 lesions. Eligible cases had to have completed both CT simulation and MRI simulation for a number of sequences outlined in the following. Information on the GTV delineations is contained in Table 1. The location of each GTV is given in reference to the Couinaud classification of liver anatomy, commonly used in radiology reporting.

MRI and CT acquisition and characteristics

The MRI was carried out using a 1.5T GE SIGNA HDxT in the radiology department. The MRI protocol included a T1 contrast-enhanced sequence called liver acquisition volume acquisition (LAVA), a noncontrast enhanced single shot fast spin echo (SSFSE) and a diffusion weighted image (DWI). The LAVA and SSFSE sequences were taken on a voluntary end expiration breath hold. The MRI, for planning purposes, is typically acquired immediately after the simulation CT with both acquired at end-expiration breath hold to improve image registration. The DWI was a respiratory-gated sequence rather than breath hold. The end phase of expiration was chosen for the gate. Due to irregularity in some patients' breathing, only 6 patients had DWI sequences.

The volume of contrast administered for the LAVA sequence was determined according to 0.1 mL/kg body

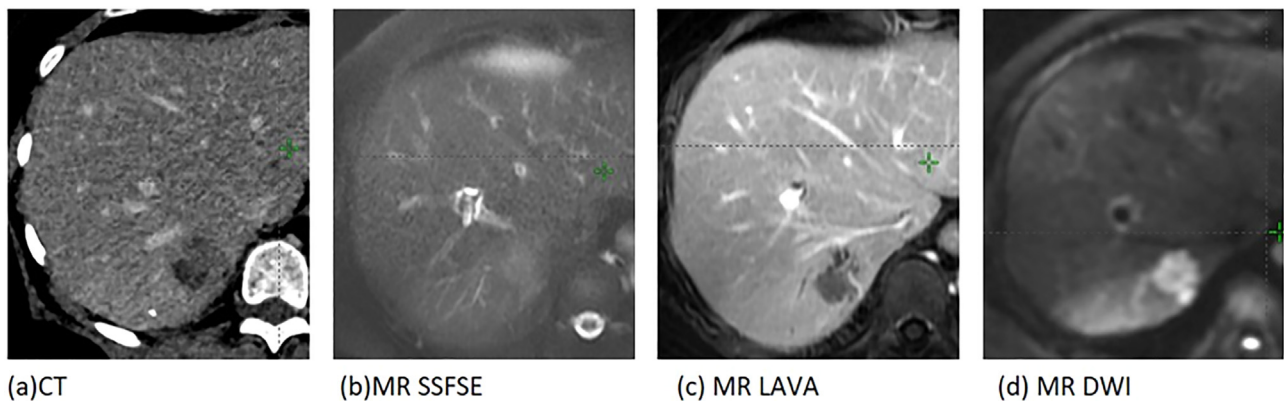


Fig. 1 The appearance of the gross tumor volume for delineation on (A) computed tomography and contrast, (B) magnetic resonance (MR) single shot fast spin echo, (C) MR liver acquisition volume acquisition, and (D) MR diffusion weighted image.

weight (0.1 mmol/kg) for each patient and images were acquired at 4 phases of contrast enhancement: (1) non-contrast, (2) arterial enhancement at 20 seconds after injection, (3) portal-venous enhancement at approximately 70 seconds after injection, and (4) a delayed contrast phase. The target appearance on a contrast enhanced T1 sequence such as LAVA includes a central hypoattenuating portion that corresponds to the central necrosis often surrounded by an ill-defined enhancing rim, which corresponds to the proliferative tumoral border. Delayed enhancement may also be present owing to the desmoplastic reaction.

The LAVA sequence is a T1 fat-saturated 3-dimensional acquisition. This is a fast sequence with the aim of acquiring the whole liver within 1 breath hold. The LAVA sequence had a slice thickness of 2.5 mm. The DWI was acquired with b values of 50 and 800. The SSFSE and the DWI sequences were low-resolution scans with slice thicknesses of 8 mm, and would not be used in isolation for GTV delineation. An example of the appearance of each image set can be seen in Fig. 1.

The CT simulation was acquired on a GE Lightspeed RT. The scans were taken at 60 seconds after contrast in end-expiration breath hold. The contrast was Omnipaque with a concentration of 70 to 80 ml and a flow rate of 1.5 to 1.7 mL/s. Contrast was not varied with patients' weight. Seven of the scans had 2.5-mm slice thickness, 2 had 5-mm slice thickness, and 1 had 1.25-mm slice thickness.

Delineations

The contouring process included 2 steps. First, each case was reviewed by a senior radiologist (>10 years of experience) who chose the most appropriate contrast-enhanced sequence for the delineation. Delineation instructions were provided for each GTV. The instructions included (1) slice visible, (2) estimate of tumor

volume dimension, and (3) appearance on the image, for example, dark in respect to surrounding parenchyma.

Contour analysis

Owing to the irregular shapes of tumors, evaluating both the overlap and the boundary differences between the GTV delineations are important.⁹ Three metrics were chosen: the conformity index, the DICE coefficient, and the maximum Hausdorff distance (HDmax).¹⁰ All analyses were conducted using SlicerRT 4.10.2.¹¹

The conformity index is the ratio of the common volume of all 3 GTVs to the encompassing volume of all 3 radiation oncologists' GTVs.¹²

The DICE coefficient is also an overlap-based metric. A pairwise comparison of each observer's delineation was performed (ie, interobserver 1 to interobserver 2, interobserver 2 to 3, and interobserver 1 to 3). The DICE ratio is the ratio of the common volume to the encompassing volume and varies from 0 (no overlap) to 1 (complete overlap).

The HDmax is a spatial distance metric that considers boundary errors in the delineation.¹⁰ The undirected is measured as the HDmax distance from boundary X to Y or from boundary Y to X. The Slicer 4.10.2 "segment comparison" module gives the undirected HDmax, which is considered in 3-dimensional form for the delineations. A pairwise HDmax was performed for each GTV delineated.

Statistical analysis

Both the conformity index and the DICE coefficient range from 0 to 1, with less interobserver variability as the metric approaches 1. The resultant data, where no manipulation of the data is carried out, is not normally distributed. A Student t test was therefore not appropriate.

The Hausdorff distance is a distance metric in which lower values demonstrate lower interobserver variability, yielding data that are not normally distributed. Thus, significance of the difference in means of the DICE, HDmax and the conformity index were analyzed using a 1-sided nonparametric permutation test, according to Ernst.¹³

In this 1-sided test, the observed data sets were resampled and the difference in the parameter to be tested (in this case the mean) of the resampled sets was calculated. As the number of combinations can be large (30 MR LAVA and 27 CT amounted to 1.4×10^{16} combinations), a Monte Carlo approach was used to evaluate n permutations. An n of 100,000 was used for the DICE and HDmax. The P value of the test is the number of combinations in which the difference in the mean is equal to or greater than the measured mean difference, divided by the number of samples.

A P value $<.05$ was considered statistically significant.

Comparison of CT and MR LAVA

The ratio of the volume of the GTV delineated by each observer on the MR LAVA and the CT was evaluated. To compare the delineations, a registration between the CT and MR was performed. A rigid registration using Eclipse version 15.5 was used to register the images in the area of the GTV. Surrounding vessels were used as a guide for the registration. Each registration was checked by a second experienced physicist, by checking the anatomy in proximity to the tumor, most commonly using vessels. In one case, where a large deformation was observed, a deformable registration was required. The Velocity 4.1 program (Varian Medical Systems) was used for deformable image registration.

Margin

The PTV in ICRU 83 is a geometric concept, whereby adding a margin on the GTV and/or clinical target volume (CTV) we are delivering a clinically accepted probability adequate dose to the GTV. All geometric uncertainties are included, including respiratory motion. Our liver SABR treatments are conducted in end-expiration breath hold, eliminating the effect of respiratory motion.

Several mathematical formulae have been recommended for generating the GTV-PTV margins. In this study, we used the van Herk recipe¹⁴ to demonstrate the difference in the margin required based on the interobserver variability seen with MR LAVA and CT. To ensure that the minimum dose of 95% to the GTV to 90% of the patients, the Van Herk margin recipe ($2.5\Sigma + 0.7\sigma$) is used, which requires a margin that is 2.5 times the total standard deviation (SD) of the systematic errors (Σ) and 0.7 times the SD of the random errors (σ).

Using the Velocity 4.1 software package, the mean distance between the boundary of the GTVs for the MR LAVA and the contrast-enhanced CT was evaluated. The package computes the mean value of the closest point from one boundary to the closest point on the second boundary volume. To determine the margin difference, 2.5 times the total SD of this boundary distance was determined.

Results

Graphical representations of the pairwise DICE similarity coefficient and the pairwise HDmax are shown in Figs. 2 and 3. The conformity index is summarized in Table 2. MR LAVA showed less interobserver variation than CT, MR SSFSE, or DWI. The overall mean DICE coefficients for MR LAVA, CT, MR SSFSE, and DWI were 0.82, 0.74, 0.55, and 0.76, respectively (Table 2). The overall mean HDmax for the MR LAVA, CT, MR SSFSE and DWI were 5.68 mm, 7.25 mm, 19.34 mm, and 9.51 mm, respectively. Similarly, the overall mean conformity indices for MR LAVA, CT, MR SSFSE, and DWI were 0.58, 0.47, 0.29, and 0.46.

For all 3 metrics, MR LAVA showed the lowest interobserver variability. CT with contrast had a slightly lower mean DICE than DWI, but the mean HDmax and mean conformity index was lower for CT with contrast. A summary of this data is available in Table 3.

As seen in Figs. 3 and 4, large variability in contouring on the noncontrast SSFSE was evident, with GTV 5 and GTV 7 having no overlap in the contouring, giving DICE values of 0. In addition, the average of the HDmax for MR SSFSE was 19.34 mm, with values ranging from 2.7 to 47 mm. From the limited number of DWI data sets, the mean DICE was slightly higher than CT at 0.76, but the HDmax (9.51 mm) and conformity index (0.46) indicated more variability in contouring.

Interobserver variability can be accounted for in the planning margin on the GTV as a systematic error. The pairwise mean distance between the boundary of the GTVs delineated on CT and MR LAVA was 1.8 mm and 1.3 mm, respectively. With an SD on the mean of 1.6 mm for CT and 1.2 mm for MR LAVA, the resulting margins, according to the Van Herk formula,¹⁴ required to account for interobserver variability would be 4 mm (CT) and 3.1 mm (MR LAVA).

Permutation test

The permutation test results are shown in Table 4. A statistically significant difference ($P < .01$) was found between the mean DICE for CT (0.74) and MR LAVA (0.82). The mean HDmax for CT (7.25 mm) and mean HDmax for MR LAVA (5.68 mm) were also found to be

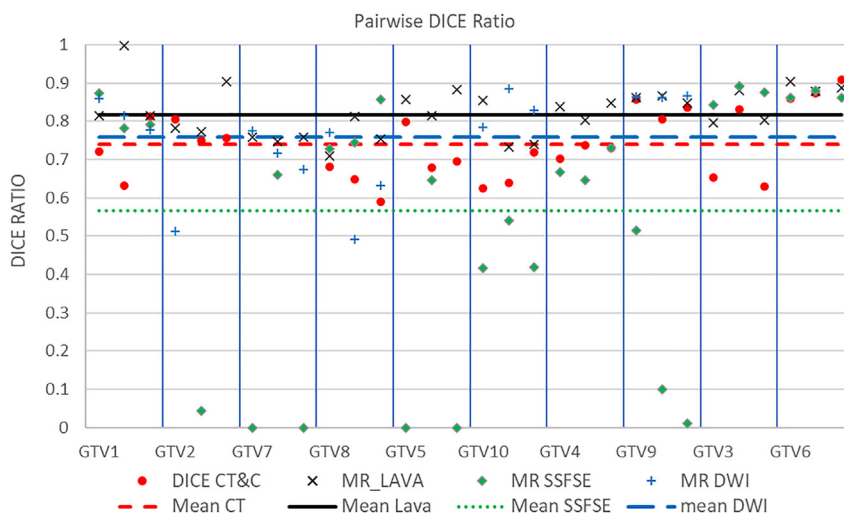


Fig. 2 Pairwise DICE ratio comparison of interobserver 1 and 2, interobserver 1 and 3, and interobserver 2 and 3 for each of the 10 GTVs in order of GTV size. Abbreviations: CT = computed tomography; CT&C = CT and contrast; DWI = diffusion weighted image; GTV = gross tumor volume; LAVA = liver acquisition volume acquisition; MR = magnetic resonance; SSFSE = single shot fast spin echo.

significantly different ($P = .04$). The difference in mean conformity index of CT (0.47) and MR LAVA (0.58) was not found to be statistically significant ($P = .08$).

Comparison of MR LAVA and CT

Figure 4 is a graphical representation of the ratio of the volume of GTV delineated on MR LAVA to CT for each

observer in order of GTV volume. Each of the observers’ GTV delineations on CT was compared with MR LAVA; 68% of volumes drawn on MR LAVA were larger than on CT ($P < .01$). By dividing the volumes into those with a value of less than 5 cc, it was shown that the effect is more significant for small GTVs. In this case, 87% of GTVs with a volume of ≤ 5 cc were smaller on CT than on MR LAVA ($P \leq 0.01$), and 53% of those > 5 cc were smaller on CT ($P = .57$). All the MR LAVA scans had 2.5-mm

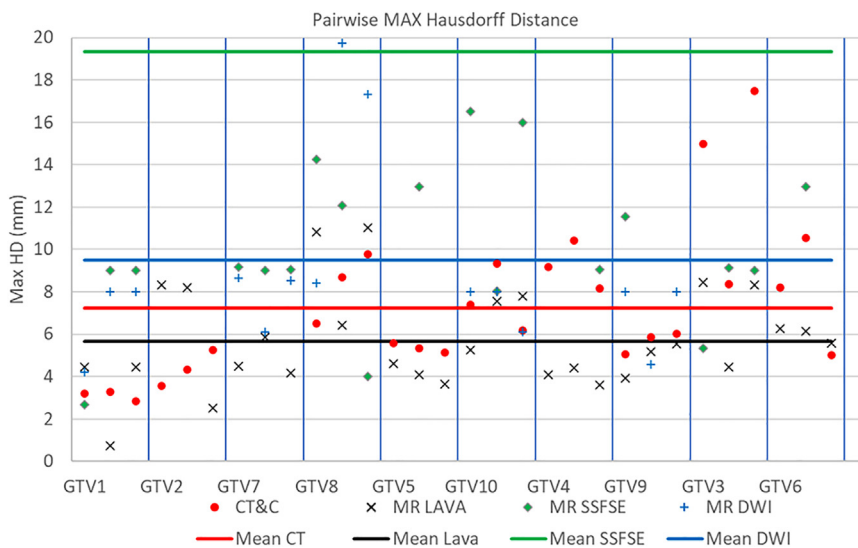


Fig. 3 Pairwise maximum Hausdorff distance of interobservers 1 and 2, interobservers 1 and 3, and interobservers 2 and 3 in order of GTV size. Large maximum Hausdorff distance values > 20 mm were seen in MR SSFSE, but they are not included in this graph. Abbreviations: CT = computed tomography; CT&C = CT and contrast; DWI = diffusion weighted image; GTV = gross tumor volume; LAVA = liver acquisition volume acquisition; MR = magnetic resonance; SSFSE = single shot fast spin echo.

Table 2 Conformity index and overlap volume of all 3 GTVs divided by the encompassing volume of all 3 GTVs for CT&C, MR LAVA, MR SSFSE, and MR DWI

| GTV | CT&C | MR LAVA | MR SSFSE | MR DWI |
|--------|---------|---------|----------|--------|
| GTV 1 | 0.42 | 0.67 | 0.45 | 0.40 |
| GTV 2 | 0.5 | 0.56 | 0.01 | 0.30 |
| GTV 3 | 0.49 | 0.65 | 0.68 | No DWI |
| GTV 4 | 0.45 | 0.64 | 0.37 | No DWI |
| GTV 5 | 0.44 | 0.56 | 0.0 | No DWI |
| GTV 6 | 0.70 | 0.74 | 0.68 | No DWI |
| GTV 7 | No CT&C | 0.48 | 0 | 0.48 |
| GTV 8 | 0.33 | 0.47 | 0.50 | 0.29 |
| GTV 9 | 0.61 | 0.67 | 0.0 | 0.65 |
| GTV 10 | 0.35 | 0.34 | 0.16 | 0.62 |

Abbreviations: CT&C = computed tomography and contrast; DWI = diffusion weighted image; GTV = gross tumor volume; LAVA = liver acquisition volume acquisition; MR = magnetic resonance; SSFSE = single shot fast spin echo.

slice thickness and 7 of the CT scans had 2.5-mm slice thickness; however, GTV 4 and GTV 5 had 5-mm slice thickness. Given the size of GTV 5, reported by radiology as 2 cm, a finer resolution along the Z axis (superior/inferior) would be appropriate.

Discussion

Interobserver variability in delineation of the GTV is a widely accepted source of uncertainty in radiation therapy and has a direct effect on the GTV to PTV margin. In this study, we examined the interobserver variability on a range of image sets with the aim of determining the most appropriate image set for GTV delineation. A secondary aim was to compare the GTVs delineated on MR to those on CT.

A thorough analysis of the interobserver variability in delineation was achieved by using a range of metrics that

consider both the overlap ratio and the boundary differences. The analysis showed MR LAVA had the lowest interobserver variability compared with CT, MR SSFSE, and MR DWI. Two of the metrics used, the HDmax and the DICE coefficient showed a statistically significant improvement in the interobserver variability on MR LAVA compared with CT.

SSFSE is a very fast imaging sequence and is used in body imaging where bowel and respiratory motion are an issue. However, this results in images with lower signal to noise, blurring and reduced image contrast. The large interobserver variability found in this study for SSFSE is not unexpected and, while useful for diagnostic purposes, this study found that the variability renders it unsuitable for use in radiation therapy as a delineation image set.

There are few studies that have examined interobserver variability of GTV delineation in the liver. One such study by Jensen et al⁷ included patients with hepatocellular carcinoma (n = 6) and metastatic liver tumors (n = 6), and the observers included 2 radiation oncologists, 2 radiation therapists, and 1 radiology resident. The volumes were delineated on a dynamic contrast-enhanced CT scan and a 4-dimensional CT scan with the analysis including the DICE coefficient but no boundary difference metrics. As such, the results presented by Jensen et al⁷ were not directly comparable to this study because it used different image sets, along with a more varied patient group and observer set.

The results of this study allow for the accurate estimate of the systematic error introduced by the interobserver variability, which is added to the margin recipe for calculation of the planning target volume (PTV). The margin adds a buffer to account for the uncertainties in the delineation of the GTV (ICRU 83⁸). This study yielded a reduction of the interobserver variability from 1.6cm (SD) for CT to 1.2cm (SD) for MR LAVA.

Steenbakkens et al¹² studied the effect on interobserver variability for lung cancer delineation using positron emission tomography (PET) CT in comparison to CT alone. The overall interobserver variability was reduced from 1 cm (SD) to 0.4 cm (SD) when using CT versus

Table 3 Comparison of CT, MR LAVA, MR SSFSE, and MR DWI mean and SD data for each metric

| | CT&C | MR LAVA | MR SSFSE | MR DWI |
|------------------|-------------|-------------|--------------|-------------|
| Metric | Mean (SD) | Mean (SD) | Mean (SD) | Mean (SD) |
| DICE | 0.74 (0.09) | 0.82 (0.06) | 0.55 (0.34) | 0.76 (0.12) |
| HDmax (mm) | 7.25 (3.45) | 5.68 (2.31) | 19.34 (15.5) | 9.51 (5.01) |
| Conformity index | 0.47 (0.12) | 0.58 (0.12) | 0.29 (0.28) | 0.46 (0.16) |

Abbreviations: CT&C = computed tomography and contrast; DWI = diffusion weighted image; HDmax = maximum Hausdorff distance; LAVA = liver acquisition volume acquisition; MR = magnetic resonance; SD = standard deviation; SSFSE = single shot fast spin echo.

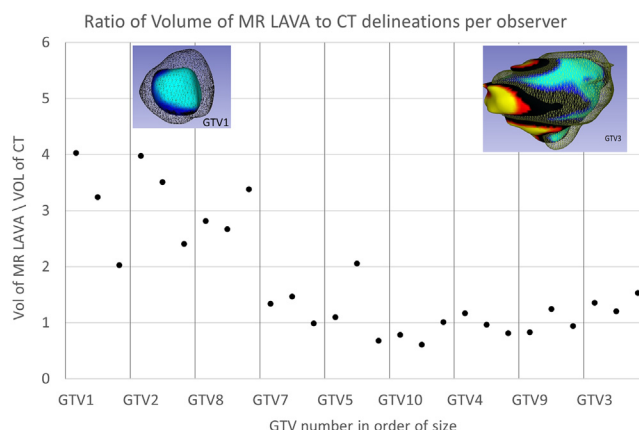


Fig. 4 Ratio of the volume of the GTV drawn on MR LAVA to CT by each observer. Inset GTV 1 and GTV 3 are 3-dimensional models. The wireframe is the MR LAVA and the solid structure is the CT. Abbreviations: CT = computed tomography; GTV = gross tumor volume; LAVA = liver acquisition volume acquisition; MR = magnetic resonance.

PET CT alone. This much lower interobserver variability in lung than liver can be expected considering the less well-defined boundaries and artifacts due to bowel and respiratory motion in liver. PET CT can be useful in highlighting a Biological Target Volume in liver SBRT. However, Riou et al,¹⁵ in their study of the benefit of 4-dimensional–PET CT in volume delineation for liver SBRT, found that nonrespiratory gated PET in the liver can result in a possible underestimation or a complete miss of the target volume.

By introducing MRI as an image set for delineation, the interobserver variability is reduced but this study also saw a significant difference in the volume of the GTV delineated on MRI in comparison to CT for small tumors. For the LAVA sequence, when GTVs delineated were 5 cc or less, the volume delineated on MRI was larger in 87% of cases, with a mean ratio of MRI volume to CT volume of 2.52. Previous studies have investigated the differences between CT and MR delineation. Pech et al¹⁶ studied 25 patients with 43 colorectal liver metastases. Similar to our study, they reported that the volume on contrast enhanced CT (mean volume, 20 mL) was less than that on the T1 weighted contrast enhanced MRI sequence

(mean volume, 65 mL). The PV phase of CT contrast enhancement was used in this study.

A limitation of these studies is the lack of literature currently available that compares imaging to histopathology. These studies are technically difficult, specifically in the preparation of the specimen. The histopathology correlation of T1 weighted images was studied by Outwater et al in 1991.¹⁷ This study reported low intensity regions corresponded to histologic findings of coagulative necrosis and desmoplasia within the tumor. The study also found that peripheral hyperintense halos around central hypointense areas encompassed the growing tumor margin and variable degrees of cell necrosis. Another matter for consideration is whether microscopic tumor beyond the macroscopic tumor can be depicted with imaging.¹⁸ Traditionally, in stereotactic radiation therapy a CTV margin for microscopic extension is not used. However, there is debate in the case of the liver, with some clinical groups adding up to an 8-mm CTV margin.¹⁹ Pech et al¹⁶ proposed that the contrast enhancing tissue is more at risk of carrying tumor cells, and by including this area on contrast enhancement on the MRI in the GTV, the CTV is included.

The American Association of Physicists in Medicine and UK SABR²⁰ consortium recommend CT and MRI for delineation of tumor volumes. We routinely employ MR imaging for tumor delineation in our clinic and, indeed, a range of MR sequences had been presented for radiation oncologist delineation until the completion of this study. With evidence from this work, the number of acquired MR sequences has been significantly reduced, eliminating the use of SSFSE in most cases while focusing on the MR LAVA sequence, which returned the lowest interobserver variability. As a result, the abridged imaging protocols have led to time savings on the MRI scanner with a resultant increased efficiency within the radiology department. Further work is required to investigate the interobserver

Table 4 Permutation test P value results of each image set mean metric value compared with magnetic resonance LAVA

| Metric | CT | SSFSE | DWI |
|------------------|------|-------|------|
| DICE | <.01 | <.01 | .01 |
| HDmax | .04 | <.01 | <.01 |
| Conformity index | .08 | .02 | .09 |

Abbreviations: CT = computed tomography; DICE = xxx; DWI = diffusion weighted image; HDmax = maximum Hausdorff distance; LAVA = liver acquisition volume acquisition; SSFSE = single shot fast spin echo.

variability when using the DWI as we had a limited number of data sets available. However, this study highlighted the potential for improvements in the MR DWI resolution, an investigation which, in collaboration with the radiology department, is ongoing.

When using MRI in conjunction with CT for treatment planning, registration of the images is required, which may introduce delineation errors, especially in the case of the liver. It is, thus, imperative to employ deformable registration. Voroney et al¹⁹ showed the need for deformable registration, demonstrating how the error can be magnified for smaller tumors in cases where the deformable registration is not used. According to American Association of Physicists in Medicine Task Group 132,²¹ an estimation of this error should be taken into account in margin recipes.

Reducing the interobserver variability in liver stereotactic radiosurgery is desirable to reduce margins and allow a therapeutic ratio necessary for tumor ablation. MR LAVA provided the lowest interobserver variability of the image sets studied. There may be a systematic error introduced for smaller tumors where MR is not used for delineation. The limited sample size of this study means that the investigation is exploratory in nature. Further work would be required to assess any systematic difference in the delineation of small tumors on MR LAVA images compared with CT. Nevertheless, studying the interobserver variability informed on the target margin necessary for accounting for such variability, and may help in determining improvements in treatment precision and standardization. The addition of automatic segmentation techniques may further assist in standardizing tumor delineation. Indeed, the recent literature indicates that there have been significant advances in tumor delineation using neural networks.^{22,23}

Conclusion

The use of MRI to complement CT in the delineation of the target in the treatment of colorectal liver metastases with SABR gives an advantage by significantly reducing the interobserver variability. The MR sequence that showed the least variability in delineation of the target was the MR LAVA.

References

- Jaffray DA. Image-guided radiotherapy: From current concept to future perspectives. *Nat Rev Clin Oncol*. 2012;9:688–699.
- Benedict SH, Yenice KM, Followill D, et al. Stereotactic body radiation therapy: The report of AAPM Task Group 101. *Med Phys*. 2010;37:4078–4101.
- Rusthoven KE, Kavanagh BD, Cardenes H, et al. Multi-institutional phase I/II trial of stereotactic body radiation therapy for liver metastases. *J Clin Oncol*. 2009;27:1572–1578.
- Palma DA, Olson R, Harrow S, et al. Stereotactic ablative radiotherapy versus standard of care palliative treatment in patients with oligometastatic cancers (SABR-COMET): A randomised, phase 2, open-label trial. *Lancet*. 2019;393:2051–2058.
- Palma DA, Olson R, Harrow S, et al. Stereotactic ablative radiotherapy for the comprehensive treatment of oligometastatic cancers: Long-term results of the SABR-COMET phase II randomized trial. *J Clin Oncol*. 2020;38:2830–2838.
- Vinod SK, Jameson MG, Min M, Holloway LC. Uncertainties in volume delineation in radiation oncology: A systematic review and recommendations for future studies. *Radiother Oncol*. 2016;121:169–179.
- Jensen NK, Mulder D, Lock M, et al. Dynamic contrast enhanced CT aiding gross tumor volume delineation of liver tumors: An interobserver variability study. *Radiother Oncol*. 2014;111:153–157.
- Hodapp N. Der ICRU-Report 83: Verordnung, dokumentation und kommunikation der fluenzmodulierten photonenstrahlentherapie (IMRT) [The ICRU Report 83: Prescribing, recording and reporting photon-beam intensity-modulated radiation therapy (IMRT)]. *Strahlenther Onkol*. 2012;188:97–99. [in German].
- Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. *BMC Med Imaging*. 2015;15:29.
- Taha AA, Hanbury A. An efficient algorithm for calculating the exact Hausdorff distance. *IEEE Trans Pattern Anal Mach Intell*. 2015;37:2153–2163.
- Pinter C, Lasso A, Wang A, Jaffray D, Fichtinger G. SlicerRT: Radiation therapy research toolkit for 3D Slicer. *Med Phys*. 2012;39:6332–6338.
- Steenbakkers RJ, Duppen JC, Fitton I, et al. Reduction of observer variation using matched CT-PET for lung cancer delineation: A three-dimensional analysis. *Int J Radiat Oncol Biol Phys*. 2006;64:435–448.
- Ernst MD. Permutation methods: A basis for exact inference. *Statist Sci*. 2004;19:676–685.
- van Herk M, Remeijer P, Rasch C, Lebesque JV. The probability of correct target dosage: Dose-population histograms for deriving treatment margins in radiotherapy. *Int J Radiat Oncol Biol Phys*. 2000;47:1121–1135.
- Riou O, Serrano B, Azria D, et al. Integrating respiratory-gated PET-based target volume delineation in liver SBRT planning, a pilot study. *Radiat Oncol*. 2014;9:127.
- Pech M, Mohnike K, Wieners G, et al. Radiotherapy of liver metastases. Comparison of target volumes and dose-volume histograms employing CT- or MRI-based treatment planning. *Strahlenther Onkol*. 2008;184:256–261.
- Outwater E, Tomaszewski JE, Daly JM, Kressel HY. Hepatic colorectal metastases: Correlation of MR imaging and pathologic appearance. *Radiology*. 1991;180:327–332.
- Okano K, Yamamoto J, Kosuge T, et al. Fibrous pseudocapsule of metastatic liver tumors from colorectal carcinoma. Clinicopathologic study of 152 first resection cases. *Cancer*. 2000;89:267–275.
- Voroney JP, Brock KK, Eccles C, Haider M, Dawson LA. Prospective comparison of computed tomography and magnetic resonance imaging for liver cancer delineation using deformable image registration. *Int J Radiat Oncol Biol Phys*. 2006;66:780–791.
- Hanna GG, Murray L, Patel R, et al. UK consensus on normal tissue dose constraints for stereotactic radiotherapy. *Clin Oncol*. 2018;30:5–14.
- Brock KK, Mutic S, McNutt TR, Li H, Kessler ML. Use of image registration and fusion algorithms and techniques in radiotherapy: Report of the AAPM Radiation Therapy Committee Task Group No. 132. *Med Phys*. 2017;44:e43–e76.
- Lundervold AS, Lundervold A. An overview of deep learning in medical imaging focusing on MRI. *Z Med Phys*. 2019;29:102–127.
- Bousabarah K, Ruge M, Brand JS, et al. Deep convolutional neural networks for automated segmentation of brain metastases trained on clinical data. *Radiat Oncol*. 2020;15:87.