UC Davis
UC Davis Previously Published Works

Title

Treatment Planning and Volumetric Response Assessment for Yttrium-90 Radioembolization: Semiautomated Determination of Liver Volume and Volume of Tumor Necrosis in Patients with Hepatic Malignancy

Permalink

https://escholarship.org/uc/item/5f70d5fk

Journal

CardioVascular and Interventional Radiology, 34(2)

ISSN

0174-1551

Authors

Monsky, Wayne L
Garza, Armando S
Kim, Isaac
et al.

Publication Date

2011-04-01

DOI

10.1007/s00270-010-9938-3

Copyright Information

Peer reviewed

CLINICAL INVESTIGATION

Treatment Planning and Volumetric Response Assessment for Yttrium-90 Radioembolization: Semiautomated Determination of Liver Volume and Volume of Tumor Necrosis in Patients with Hepatic Malignancy

Wayne L. Monsky · Armando S. Garza · Isaac Kim · Shaun Loh · Tzu-Chun Lin · Chin-Shang Li · Jerron Fisher · Parmbir Sandhu · Vishal Sidhar · Abhijit J. Chaudhari · Frank Lin · Larry-Stuart Deutsch · Ramsey D. Badawi

Abstract

Purpose The primary purpose of this study was to demonstrate intraobserver/interobserver reproducibility for novel semiautomated measurements of hepatic volume used for Yttrium-90 dose calculations as well as whole-liver and necrotic-liver (hypodense/nonenhancing) tumor volume after radioembolization. The secondary aim was to provide initial comparisons of tumor volumetric measurements with linear measurements, as defined by Response Evaluation Criteria in Solid Tumors criteria, and survival outcomes.

Methods Between 2006 and 2009, 23 consecutive radioembolization procedures were performed for 14 cases of hepatocellular carcinoma and 9 cases of hepatic metastases. Baseline and follow-up computed tomography obtained 1 month after treatment were retrospectively analyzed. Three observers measured liver, whole-tumor, and tumor-necrosis volumes twice using semiautomated software.

Results Good intraobserver/interobserver reproducibility was demonstrated (intraclass correlation [ICC] > 0.9) for tumor and liver volumes. Semiautomated measurements of liver volumes were statistically similar to those obtained with manual tracing (ICC = 0.868), but they required significantly less time to perform ($p < 0.0001$, ICC = 0.088). There was a positive association between change in linear tumor measurements and whole-tumor volume ($p < 0.0001$). However, linear measurements did not correlate with volume of necrosis ($p > 0.05$). Dose, change in tumor diameters, tumor volume, and necrotic volume did not correlate with survival ($p > 0.05$ in all instances). However, Kaplan–Meier curves suggest that a >10% increase in necrotic volume correlated with survival ($p = 0.0472$).

Conclusion Semiautomated volumetric analysis of liver, whole-tumor, and tumor-necrosis volume can be performed with good intraobserver/interobserver reproducibility. In this small retrospective study, measurements of tumor necrosis were suggested to correlate with survival.

Keywords Radioembolization · CT · Volume · Tumor · Response · Necrosis

W. L. Monsky (✉) · S. Loh · J. Fisher · P. Sandhu · V. Sidhar · A. J. Chaudhari · F. Lin · L.-S. Deutsch · R. D. Badawi
Department of Radiology, University of California Davis Medical Center, 4860 Y Street, Ste. 3100, Sacramento, CA 95817, USA
e-mail: wayne.monsky@ucdmc.ucdavis.edu

A. S. Garza · I. Kim
University of California Davis Medical School, Sacramento, CA 95817, USA

T.-C. Lin
Department of Statistics, University of California Davis, Davis, CA 95616, USA

C.-S. Li
Division of Biostatistics, Department of Public Health Services, University of California Davis, Davis, CA 95616, USA

Introduction

Dose calculations for radioembolization using Yttrium-90 (^{90}Y) microspheres, i.e., Theraspheres (MDS Nordion, Ottawa, Ontario, Canada) and Sirspheres (Sirtex Medical, Wilmington, MA), require calculation of liver volume and estimation of tumor burden [1]. A number of techniques have been described to determine liver volume for planning liver transplant. Volumetric measurements of tumors and organs, as seen on computed tomography (CT) images, are

often performed by manual tracing of the organ boundary and summation of the liver area on each section, taking into account the thickness of each image slice. However, the manual tracing process is time consuming [2, 3] and arguably unreliable and inaccurate because it is highly operator dependent. Techniques for the automated determination of liver volumes could aid in planning the ^{90}Y dose for the treatment of hepatic tumors.

Furthermore, response to these local regional therapies can be assessed using semiautomated segmentation. The technique can be used to determine the change in volume of necrotic and viable tumor. However, these techniques must first be validated, and interobserver/intraobserver reproducibility must be demonstrated. These volumetric analyses should be compared with Response Evaluation Criteria in Solid Tumors (RECIST) criteria, the current standard for tumor response assessment, to suggest clinical applicability. Moreover, to be clinically useful, tumor response analysis should provide an early indicator of survival outcomes.

In the recent RECIST 1.1 criteria, it was recognized that volumetric techniques may improve on the current standard of single long-axis tumor measurements, which may be inadequate for evaluating tumor response after biologic, cytostatic, and local regional therapies [4]. Tumor-response measurements described by RECIST criteria evaluate only unidimensional tumor measurements and disregard the extent of necrosis, which includes the initial effect of all local regional therapies. In one study, RECIST criteria missed all complete responses and underestimated the extent of partial tumor response because of tissue necrosis, thus incorrectly assessing the therapeutic efficacy of local regional therapy [5].

European Association for the Study of the Liver (EASL) guidelines recommended that assessment of tumor response should incorporate the decrease in viable tumor burden and that tumor necrosis should be measured when evaluating therapeutic response in hepatic tumors. These measurements are usually based on linear measurements of the tumor and necrosis. Contrast-enhanced and diffusion-weighted magnetic resonance imaging (MRI) has been used to evaluate tumor necrosis after catheter-directed local regional therapies [6]. Tumor necrosis is often patchy and irregular, making linear measurements difficult and more susceptible to observer-based error as well as poor measurement reproducibility. The computer-assisted semiautomated segmentation and volumetric analysis techniques we describe here may improve measurement reproducibility, particularly when measuring complex and patchy tumor necrosis and enhancement. This volumetric analysis of tumor necrosis could provide an earlier indication of therapeutic efficacy.

Volumetric tumor response has been applied to lung tumors and lymphoma [7]. Few studies have evaluated volumetric techniques for the assessment of hepatic tumors [8]. Volumetric segmentation techniques have not been applied to response assessment after local regional therapies, such as radioembolization and chemoembolization. The purpose of this study was first to demonstrate good interobserver/intraobserver reproducibility when measuring liver volume for treatment planning as well as when measuring viable and necrotic, hypodense tumor after local regional therapy. Changes in whole-tumor and necrotic-tumor volume were then compared with linear measurements, as described by RECIST criteria, to determine any association. Finally, in this small pilot study, these measurements were compared with patient survival to provide proof of the principle that they might be used as early imaging end points for tumor-response assessment.

Materials and Methods

Subject Demographics

The study was a cohort study in which volumetric analysis techniques were applied in retrospective fashion to CT images obtained during the course of previous radioembolization treatments. Institutional Review Board (IRB) approval for the study was obtained. Data collection was performed in accordance with Health Insurance Portability and Accountability Act regulations.

Subjects were consecutive patients who had undergone radioembolization for HCC as well as hepatic metastases between 2006 and 2009 according to a protocol approved by our IRB. The study population comprised 18 men (mean age 66.1 years [range 48–79]) and 5 women (mean age 67.3 years [range 60–73]). Fourteen patients had a known diagnosis of hepatocellular carcinoma. The remaining patients had known metastasis to the liver, of which metastatic colon carcinoma was the most prevalent (5 cases), followed by metastatic carcinoid (2 cases), metastatic insulinoma (1 case), and metastatic leiomyosarcoma (1 case). The number of tumors per individual patient ranged between 3 and 12. Long-axis measurements for the tumors ranged between 1.2 and 18 cm (mean 9.8 [SD 6.9]). If tumors occupied both the right and left lobes of the liver, lobar treatment was performed with separate procedures approximately 1 month apart. In 2 patients, both lobes were treated twice. In these cases, "baseline" measurements were made using the CT image obtained within 1 month before a given treatment.

All patients underwent pretreatment assessment consisting of a clinical history, appropriate laboratory workup, and baseline four-phase liver CT image obtained within 1 month of treatment. The selection of ^{90}Y over other treatments was based on the consensus of a multidisciplinary team during tumor board conferences at our institution and in accordance with our inclusion and exclusion criteria (Table 1). No systemic therapy was administered

Table 1 Inclusion and exclusion criteria for radioembolization

Inclusion criteria

  Absence of extrahepatic cancer involvement

  Child-Pugh class A or B

  Performance status: ECOG 0–2, target liver lesion measurable per RECIST criteria, life expectancy at least 3 months in absence of treatment, creatinine $\leq$ 1.50 mg/dl, bilirubin $\leq$ 2.5 mg/dl, albumin $\geq$ 30 g/l, white blood cells $\geq$ 1.5 $\times$ 109/l, PLT $\geq$ 50 $\times$ 109/l

Exclusion criteria

  Ascites

  Contraindications to arteriography

  Presence of localized or systemic infections

  Pregnant (women of child-bearing potential will have a pregnancy test done)

  Known or suspect hypersensitivity to the chemotherapeutics

PLT platelets

once radioembolization was initiated. Sixteen of these patients had previously progressed with first- or second-line chemotherapy. Patients were seen in the follow-up clinic at 1 week, 1 month, and then at 3-month intervals while they underwent local regional therapy.

Radioembolization Procedure

The technical details for radioembolization have previously been described [1]. All patients underwent mapping angiography before treatment to determine vascular anatomy and arterial variants. Prophylactic embolization of nontarget vessels was performed to avoid nontarget deposition of microspheres. ^{99}Tc-macroaggregated albumin was used to assess lung shunt fraction and splanchnic shunting.

The ^{90}Y radioembolization dose administered to the target site was calculated using the following formula (Eq. 1):

$$D \text{ (Gy)} = [A \text{ (GBq)} \times 50(1 - \text{LSF}) \times (1 - -R)]/M \text{ (kg)},$$
$$(1)$$

where D is the dose administered to the target site, A is the actual activity delivered to the site, LSF is the lung shunt fraction, R is the residual activity in the vial, and M is the mass of the area infused [1]. The body surface area method of dose calculation, which requires calculations of the volume of the liver and the tumors to be treated, can also be used [1]. The calculated dose was based on the liver volume and percentage of the lobe being treated that was occupied by tumor. For these determinations, baseline CT obtained within 1 month of treatment was initially analyzed volumetrically. If repeated treatments were to be performed, the dose calculations were based on follow-up CT obtained within 1 month before a given treatment. The ordered and administered doses were recorded and

retrospectively tabulated for comparison with the other measured indices.

Multiphasic Abdominal CT

Details of the CT scan protocols are listed in Table 2. Baseline four-phase (nonenhanced, arterial, portal venous, and delayed phase) contrast-enhanced CT images were obtained within 1 month before radioembolization. Four-phase, contrast-enhanced CT image was also acquired 1 month after treatment to assess response to therapy and then every 3 months while the patient was undergoing radioembolization treatments. To determine if these volumetric measurements might be used for early prediction of response, we choose to evaluate volumes from CT images obtained 1 month after treatment. Digital imaging and communications in medicine (DICOM)-format images obtained for baseline and 1-month follow-up CT images were then used for retrospective application of our semi-automated segmentation software for the volumetric analysis of whole-tumor and tumor-necrosis measurements. All images were deidentified in accordance with HIPA requirements; thus, the observers were blinded to patient demographics.

Table 2 Technique for multiphasic abdominal CT

Four-phase enhanced CT for evaluation tumor response

Arterial phase

  Pitch: 1.375

  Collimation: 20 mm

  kV: 120

  mA noise index of 14 (mA range 100–440)

  125 cc nonionic contrast injected at 4 cc/s

  Scan timing: Aortic contrast arrival + 8 s (smart prep used with evaluation of the abdominal aorta)

Portal venous phase

  Pitch: 1.375

  Collimation: 20 mm

  kV: 120

  mA: Noise index of 12 (mA range 100–440

  Scan timing: 35 s after arterial phase

Equilibrium phase

  Pitch 1.375

  Collimation: 20 mm

  kV: 120

  mA: Auto mA noise index of 16 (mA range 100–440)

  Scan timing: 90 s after arterial phase

  Matrix size of 512 $\times$ 512 pixels

  Window parameters were set to the "abdomen preset" with window level = 25 HU and a window width = 175 HU

CT scans were acquired on one of two clinical CT imaging scanners (Lightspeed 16; GE Medical Systems, Milwaukee, Wisconsin). The arterial phase of enhancement was used for volumetric analysis of hepatocellular carcinoma, and the portal phase was used for metastases because it allowed for the greatest conspicuity. All CT images obtained during the care of the patient were initially stored on a Philips iSite Picture Archiving and Communication System (PACS, Philips Healthcare, Andover, MA).

Response Assessment

Response according to tumor size was initially assessed using RECIST criteria, in which the sum of the longest axial measurement of the target lesions after treatment was compared with that of the lesions at baseline [4]. When evaluating correlations of RECIST-based imaging end points with clinical end points, we chose to combine stable disease and partial response, i.e., <20% increase in tumor size, for Kaplan–Meier analysis. Tumor response was also assessed according to EASL criteria, which quantifies the amount of enhancing and necrotic nonenhancing tissue in the treated lesion [9]. However, the volume of necrotic tissue was determined using the segmentation analysis described later in the text rather than using linear measurements. When evaluating correlations of EASL-based imaging end points with survival end points, we choose a cut-off point of 10% necrosis in this pilot study. In this study, tumor necrosis was defined as hypodense nonenhancing portions of the tumor as demonstrated on contrast-enhanced CT image.

Semiautomated Segmentation and Volume Measurement

All operators were initially trained to use the software and were supervised by an attending radiologist with 10 years of experience (W. M.). A set of 10 images, not included for the actual study, and a scripted instructional session were used for training purposes.

For the assessment of interobserver/intraobserver reproducibility to determine liver volumes using the semiautomated segmentation based on iterative watershed segmentation (IWS), 23 enhanced CT images were retrospectively analyzed. Three operators (A. G., S. L., and I. K.) analyzed 23 image sets twice, separated by 1 month's time, resulting in a total of 138 liver measurements. One operator was a radiology resident (S. L.), and two operators were senior medical students (I. K. and A. G.). The operators recorded the volume in cubic millimeters along with the time elapsed in seconds for each liver measurement.

The liver volumes were also manually measured from the same CT images on our PACS system using Voxar by tracing the edge of the liver on each axial slice and then summating the areas, as currently performed clinically. Measurements made using the semiautomated segmentation software and Voxar-based manual tracing techniques were also compared. The time required to make these measurements was compared. Voxar is a commercially available (Barco, Brussels Belgium), FDA-approved, three-dimensional segmentation software incorporated in our Stentor iSite (Stentor, San Francisco, CA) PACS system. The segmentation software (www.pasa.nhs.uk/pasa/doc) uses "a number of proprietary techniques" to segment the volume of the selected tissue of interest.

To determine intraobserver/interobserver reproducibility when measuring whole-tumor, viable-tumor, and necrotic-tumor volume, three operators (W. M., S. L., and I. K.) analyzed 23 image sets twice, separated by 1 month's time, resulting in a total of 138 tumor measurements.

Long-axis measurements of these tumors were also determined, as described by RECIST criteria, by two radiology residents (P. S. and J. F.) and compared with whole-tumor and necrotic-tumor volumes as well as survival. Survival statistics were obtained from the electronic medical record (EMR) (V. S.). Twenty-one patients who had gone radioembolization were included in the assessment for correlation between volumetric and linear measurements as well as survival outcomes. Two patients were not included because they were lost to follow-up; they moved residence shortly after radioembolization. All observers were blinded to patient identity because images were deidentified, and observers recording survival times were blinded to tumor measurements. When evaluating responses, making comparisons with survival outcomes, and comparing tumor measurements, we analyzed enhanced baseline CT images obtained within 1 month before each radioembolization and follow-up enhanced CT images obtained 1 month after each radioembolization. Survival was defined as the date of diagnosis to the date of death and from the date of the first radioembolization treatment to the date of death as obtained retrospectively from the EMR.

Semiautomated Measurement of Hepatic and Tumor Volumes

Our semiautomated segmentation software was developed on a Windows XP (Microsoft Corporation, Redmond, WA) workstation with the MatLab programming language (version R2007a; The MathWorks, Natick, MA). The program was run on an HP Pavilion dv1000 Laptop with a 1.5-GHz processor, 2 GB RAM, and an 80-GB hard drive (Hewlett Packard, Palo Alto, CA). A graphical user interface was implemented. The semiautomated segmentation
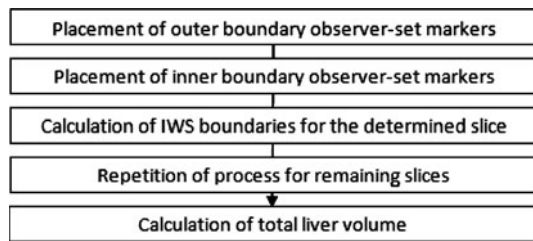
Fig. 1 Overview of IWS-based semiautomated measurement of liver volumes

software developed for volumetric analysis used algorithms based on IWS [8] and pixel thresholding (PT) [10]. The IWS algorithm implemented first used gradient vector flow transformation of the image for edge enhancement, which was followed by a region-growing method, analogous to "flooding" the gradient vector flow map. Liver and tumor edges were finally determined from subsequent ridgelines formed around each basin (local minima). This method was previously described and validated by Ray et al. [8].

The overall scheme for semiautomated volumetric measurement of the liver is illustrated in Fig. 1. All of the image sets were downloaded from PACS onto a personal computer using the DICOM format. Slices that did not include liver parenchyma were deleted from the image sets. As such, each operator performed measurements on the same set of CT images. Hepatic segmentation was performed by placing a series of markers that defined the outside of the liver as well as series of markers that defined a representative sample of the liver parenchyma. The

software automatically connected each marker with a straight line, creating a polygonal shape. On each axial image, the operators had the option of checking their work by way of an output image (Fig. 2), or to continue to the next set of images. The operators had the ability to delete, add, or move any marker at any time. When all image markers were set, the program added the individual slice volumes, and total volume was recorded. The time from loading DICOM images onto the program until output of total volume was recorded.

Semiautomated Measurement of Whole-Tumor, Viable-Enhancing Tumor, and Necrotic-Nonenhancing Tumor Volumes

Segmentation of whole tumor volume was performed by placing a series of markers that defined the liver parenchyma outside of the tumor as well as a series of markers that defined a representative sample of the tumor. The software automatically connected each marker with a straight line, creating a polygonal shape for initialization of the segmentation routine (Fig. 3A). On each axial image, the operators had the option of checking their work by way of an output image or continuing to the next set of images. The operators had the ability to delete, add, or move any marker at any time. When all image markers were set and the boundaries determined, the program added the individual slice volumes, and total volume was recorded.

The IWS algorithm calculated the whole-tumor volume (enhancing and nonenhancing) (Fig. 3A–C). This was

Fig. 2 (Left) Enhanced axial CT image through the abdomen with observer-set markers defining the outside of the liver and internal markers defining a representative sample of liver parenchyma. (Right) Image output post-IWS processing shows the estimated perimeter of the liver
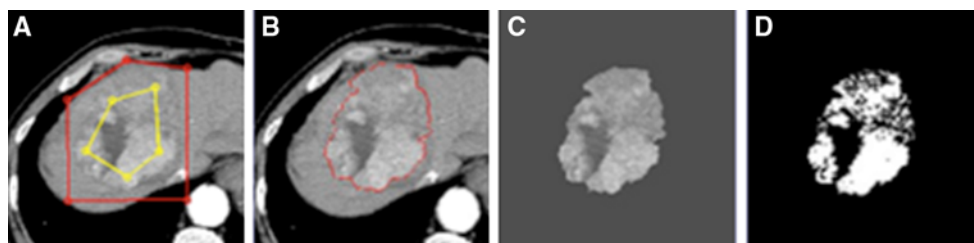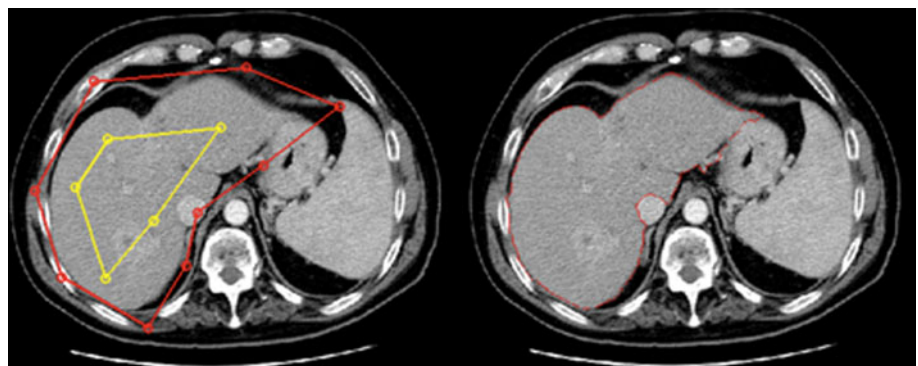
Fig. 3 A IWS operator sets internal and external markers results in B outline of whole tumor edge. C Image mask of whole tumor results in D threshold image to distinguish enhancing tumor from necrosis

followed by application of the PT algorithm, in which a region of normal parenchyma and the enhancing portion or necrotic nonenhancing portion of the tumor were included in separate regions of interest. The histogram of pixel attenuation value was evaluated. The formula for selecting pixel attenuation values was as follows (Eq. 2):

$$x = \text{mean} + (\text{std\_dev} * \text{multiplier factor}), \quad (2)$$

where x is the cut-off pixel value. All pixel values $>x$ are counted as enhancing viable tumor, mean is the average pixel value from a sample region of normal adjacent background parenchyma, and std_dev is the SD from the same sample. If we assumed normal/Gaussian distribution of the pixel values of normal parenchyma, and we took all pixels with values $>x$ multiplied by a factor of ≥ 3, 99% of the pixel attenuation values selected were higher (or more dense) than adjacent normal parenchyma. In practice, the normal parenchyma samples are not in perfect normal/Gaussian distribution. The multiplier factor, which best distinguishes enhancing viable tumor from necrosis, was chosen by the operator. Necrosis volume = whole tumor volume − viable enhanced tumor volume (Fig. 3C, D). The percentage change of each volume was obtained by comparing volumes measured on pre-radioembolization baseline CT image with CT image obtained 1 month after radioembolization. Thus, a negative % change represents reduction in volume. Target lesions that underwent radioembolization were identified and measured; untreated lesions were not included in this analysis because treatment was usually lobar.

Statistical Analysis

Intraobserver/Interobserver Reproducibility of Liver Volumes

To determine intraoperator reproducibility of measurements of whole-liver volumes using the IWS-based semiautomated method, we used two-sided Wilcoxon signed-rank test to assess whether the population median of the difference in whole-liver volume measurement between the two trials approached zero.

For interoperator reproducibility of the described volumetric method, we computed the ICC among three operators [11]. We also constructed a Bland–Altman plot [12] between operators, in which the graphs were assigned the mean of the two liver volumes, as measured by the two operators, as the abscissa (x-axis) value and the difference between the two liver volumes as the ordinate (y-axis) value.

The time to complete each measurement was recorded for each method. Friedman's test was used to compare the time required for each volumetric measurement among the three operators.

Intraobserver/Interobserver Reproducibility of Whole-, Viable-, and Necrotic-Tumor Volume Measurements

For intraoperator reproducibility of measurements, we used two-sided Wilcoxon signed-rank test to assess whether the population median of the differences in each volumetric measurement between the first and second measurements approached zero.

To determine interoperator reproducibility for each volumetric measurement, we computed the ICC among three operators [11]. We also constructed a Bland–Altman plot for each volumetric measurement between two operators, in which the graph assigned the mean of measurements, as measured by the two operators, as the abscissa (x-axis) value and the difference between the two liver volumes as the ordinate (y-axis) value (Fig. 3) [12].

Association of Delivered ^{90}Y Radiation Dose and Change in Whole-Tumor Volume, Tumor-Necrosis Volume, and Linear Tumor Measurements

We used linear regression with MM (maximum likelihood type) estimation, a combination of high breakdown value estimation and efficient estimation introduced by Yohai [13], to study the associations between the delivered dose, change in necrosis, change in linear measurements, and change in whole tumor.

Correlation of Imaging-Based End Points with Survival

To explore whether each of the variables, i.e., radiation dose change in necrosis, change in RECIST score, and change in whole tumor, was statistically significantly associated with the survival time, we used the Cox proportional hazards model [14].

Two-sided log-rank test was used to compare the survival distributions, i.e., since diagnosis and since treatment, between the group of patients with delivered dose, the group of patients with <10% versus ≥10% change in necrotic tumor volume, the group of patients with change of summated long-axis measurements <20% or ≥20%, and the group of patients with change of whole tumor volume <20% versus ≥20%. In addition, Kaplan–Meier survival curves were calculated [15].

Results

Hepatic Volume

Intraoperator Reproducibility

The average measured liver volume using the semiautomated segmentation was 2481.92 ± 1662.98 cc (Table 3).

Table 3 Whole-liver volumes determined with IWS versus Voxar

Variable	N	Mean	SD	SE	Median	Min	Max
Volume of whole liver measured by IWS	23	2481.9	1662.9	354.5	1913.2	1312.1	7577
Volume of whole liver measured by Voxar	23	2419.7	1596.4	340.3	1804.7	1311.32	7148.2

Intraoperator reproducibility was defined as the difference between the first and second volume measurements (in cubic centimeters). Measurements from each of the three observers were analyzed separately. For each observer, two-sided Wilcoxon signed-rank test was performed to determine if the population median of the volume difference approached zero. Observers no. 2 and no. 3 had p-values of 0.24 and 0.30, respectively, indicating that the median difference between trials was not different from zero. Observer no. 1 had a p-value of 0.046, indicating that there was a difference between volume measurements for the two trials that was barely statistically significant.

Interoperator Reproducibility

For each measurement, the ICC [11] among the three observers was computed. For the first and second measurements, the values were 0.984 and 0.992, respectively. Both of these values were high, indicating that the volume measurements for a given image were consistent across all three observers. The higher ICC value for the second measurement indicated that the three observers' second set of measurements was more consistent than the first set of measurements, suggesting some learning curve. The Bland–Altman [12] plots between any two operators also suggested interoperator reproducibility (Fig. 4).

There was no statistically significant change in speed of acquisition when making measurements the first and second time with either the automated or manual approach (Table 4). For each observer, two-sided Wilcoxon signed-rank test was performed to determine if the population median of the time difference between measurements approached zero ($p = 0.628$, 0.1000, and 0.427 for observers no. 1 through 3, respectively).

Comparison of Manual Versus Semiautomated Segmentation of Liver Volume

In addition to determining interobserver/intraobserver reproducibility of measurements using this semiautomated segmentation, the segmentation software was compared with our current standard manual tracing tools using the Voxar work station (Table 3). Measurements of liver volume, as well as the time required to complete the measurements, were evaluated by calculating the ICC [11] between the two techniques. An ICC of 0.868 indicated

that the measurements of whole liver by the two techniques were similar. Bland–Altman [12] plot between the two techniques also suggested similarity of liver volume (Fig. 5).

Significantly less time was required to take volumetric liver measurements using the semiautomated software compared with current manual tracing of the edge of the liver on each axial CT image (p-value of two-sided Wilcoxon signed-rank test <0.0001) (Table 4). An ICC of 0.088 also indicated a significant difference in the time required to measure whole-liver volumes using the two techniques [11]. Bland–Altman [12] plots between the two techniques also demonstrated differences in the time required to make these measurements (Fig. 6).

Intraobserver/Interobserver Reproducibility for Measurements of Whole-Tumor and Necrotic-Tumor Volumes

There was good reproducibility of volumetric measurements of whole-tumor and necrotic-tumor volume applied to tumors that had undergone radioembolization, with no significant difference between the first and second measurements within any of the three observers ($p > 0.05$). However, one observer demonstrated a statistically significant difference between the two measurements of viable enhancing tumor ($p = 0.004$).

To determine interobserver reproducibility for each volumetric measurement, the ICC among the three observers was computed [11]. Furthermore, Bland–Altman plots between any two operators were evaluated [12]. The ICC among the three observers for the first and second measurements of whole tumor volume was 0.988 and 0.992, respectively. The ICC among the three observers for the first and second measurements of the volume of non-enhancing hypodense tumor necrosis was 0.912 and 0.951, respectively. The ICC among the three observers for the first and second measurement of the volume of viable enhancing tumor was 0.899 and 0.911, respectively.

Comparisons of Linear and Volumetric Tumor Measurements

There was a statistically significant positive association between the change in linear measurements of tumors, as defined by RECIST criteria (as an independent variable),

Fig. 4 Bland–Altman plots for interoperator reproducibility. Each point represents a single patient image set, plotting on the *x*-axis (liver volume measured by operator A and liver volume measured by operator B versus the *y*-axis (liver volumetric measured by operator A and liver volumetric measured by operator B). The degree of dispersion around the horizontal axis represents variability in the measurements. Liver volumes are in cubic centimeters
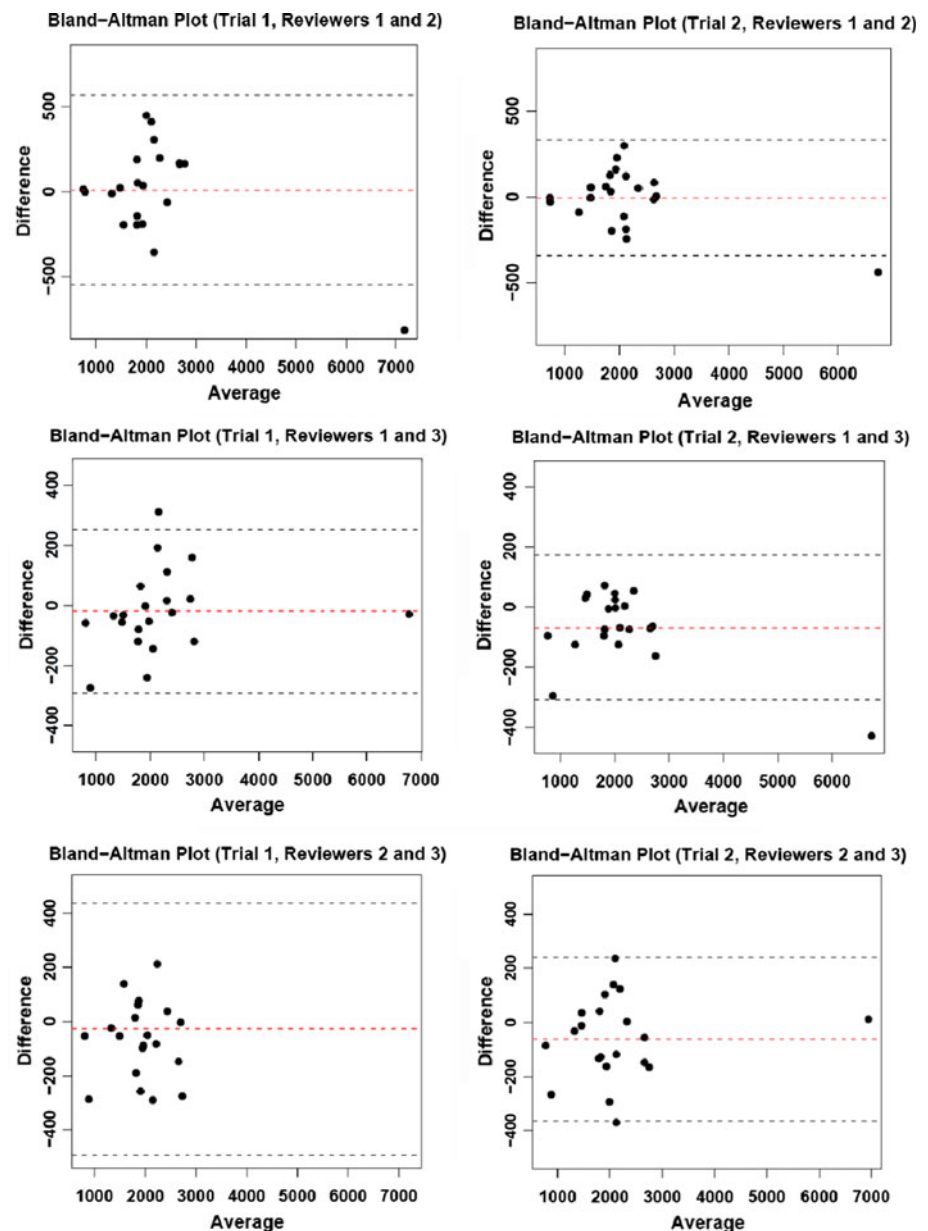
and change in whole-tumor volume (as a dependent variable) ($p < 0.0001$). However, these linear tumor measurements did not correlate with the volume of necrotic-tumor tissue ($p > 0.05$).

Correlation of Radioembolization Dose,
*Tumor Response, and Survival*

Comparisons of measurements were made between baseline CT image obtained within 1 month before a given treatment and follow-up CT image obtained 1 month after

a treatment to determine if these measurements could be used for early prediction of response and survival. A summary of the change in necrotic tumor volume, linear tumor measurements, and whole-tumor volumes, as well as delivered ^{90}Y dose, is listed in Table 5. In this small pilot study aimed at evaluating observer reproducibility for this volumetric analysis tool, we also attempted to explore the possible clinical utility of this approach.

[90]Y radiation dose delivered during radioembolization did not correlate with subsequent change in tumor measurements, including linear measurements ($p = 0.3025$),

Table 4 Time for whole-liver volume measurements with IWS versus Voxar

Variable	N	Mean	SD	SE	Median	Min	Max
Time to measure whole liver by IWS (s)	23	395	122.82	26.185	394	246	745
Time to measure whole liver by Voxar (s)	23	560.27	118.35	25.233	531.5	411	832
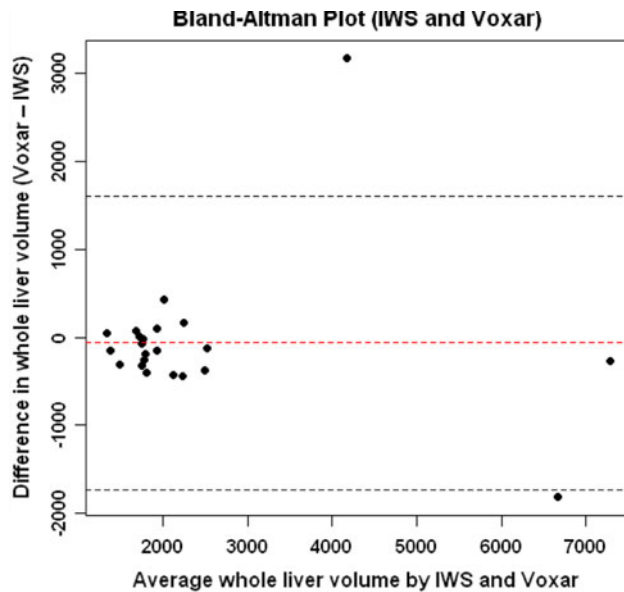
Fig. 5 Bland–Altman plot comparing measurements of whole-liver volumes using semiautomated IWS segmentation versus manual edge tracing using Voxar
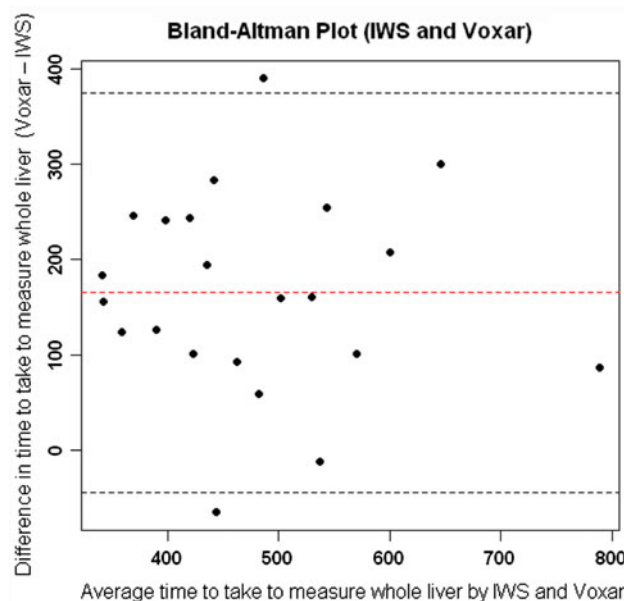
Fig. 6 Bland–Altman plot comparing time taken to make measurements of whole-liver volumes using semiautomated IWS segmentation versus manual edge tracing using Voxar

whole-tumor volume ($p = 0.4041$), and tumor-necrosis volume ($p = 0.6743$). However, such is not the aim of this article. Our patient population is likely too small for this determination. Response and survival is dependent on a large number of factors. For example, survival can be affected when there is a difference in dose administration between tumor and normal (partition model dosimetry).

Using the Cox proportional hazards model, we determined that administered ^{90}Y dose, overall change in single long-axis diameter measurements, whole-tumor volume, and necrotic-tumor volume did not correlate with survival outcomes as measured from time of diagnosis or from time of treatment ($p > 0.05$ in all instances). However, when constructing Kaplan–Meier curves, there was a statistically significant difference in survival time since diagnosis between the group of patients with change in necrosis <10% compared with those with change in necrosis ≥10% (p-value of log-rank test = 0.0472) (Fig. 7). Finally, there was no statistically significant difference in survival times when comparing the group of patients with change of summated long-axis measurements <20% or ≥20% and those with change of whole tumor volume <20% versus ≥20% (p-value of log-rank test = 0.3198 and 0.4506, respectively).

Discussion

There exists an extensive body of evidence demonstrating the palliative and therapeutic efficacy of local regional catheter-directed therapies, such as radioembolization, for the treatment of hepatocellular carcinoma as well as hepatic metastases [1]. The optimal approach for imaging-based follow-up after radioembolization is a subject of much ongoing research [5, 16]. The current study describes a technique for semiautomated volumetric assessment of the whole tumor as well as the viable enhancing and nonenhancing necrotic or nonperfused portions of the tumor. Good reproducibility of each of these measurements was demonstrated. This type of analysis might allow early prediction of response.

Furthermore, radioembolization preprocedural planning and dose determination is dependent on calculations of liver volume and the percentage of liver occupied by tumor. Organ and tumor volumetric measurements, as seen on CT images, are usually performed by manual tracing of

Table 5 Summary of tumor measurements

Variable	N	Mean	SD	SE	Median	Min	Max
[90]Y Dose (mCi)	23	36.7	15.3	3.3	32.8	16.1	66.7
Change in necrotic volume (%)	23	−31.2	77.2	16.8	−26.131	−246.17	45.2473
Change in linear measurements (%)	23	3.223	27.9	6.1	−7.5471	−30.882	94.7619
Change in whole tumor volume (%)	23	41.93	135.72	29.6	−19.894	−98.164	402.048
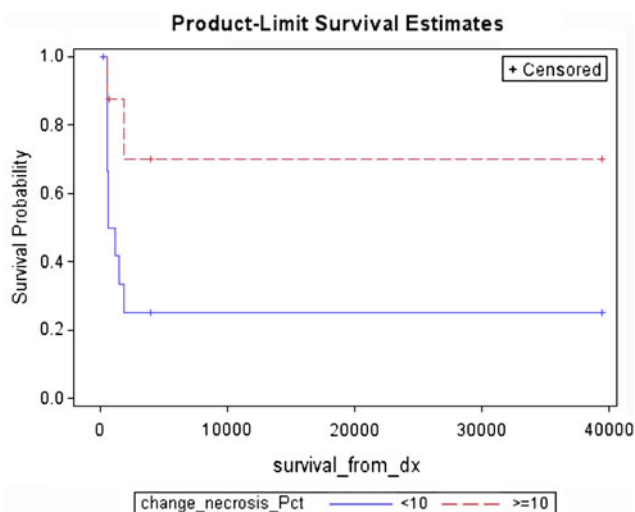
Fig. 7 Kaplan–Meier curves of survival time since diagnosis for the group of patients with change in necrosis <10% and the group of patients with change in necrosis ≥10%

boundaries and summation of the areas on each section. These manual methods of liver-volume segmentation are time consuming and laborious. They have been reported to take anywhere from 19.0 to 46.5 min [2, 3] In our study, the IWS-based semiautomated method required an average of 6.0 min across all observers, 3.2- to 7.7-fold faster than reported manual methods. We found manual-based tracing using Voxar to take significantly longer to complete. In practical terms, the method represents an improvement compared with manual methods because it only requires the approximate creation of a rough polygon bordering around a tissue of interest instead of a labor-intensive "hand-drawn" accurate border.

It has been suggested that measuring tumor response with linear diameters alone may be limited because tumor necrosis is not taken into account. [9]. This is particularly true when evaluating response to local regional therapies, such as radioembolization [17]. Characteristics of tumor enhancement after radioembolization were studied by Riaz et al. [18] and Ibrahim et al. [19]. Enhancement characteristics were not included in the current study. However, tumor necrosis is often patchy and irregular in distribution, making usual linear measurements difficult. In addition, response to local regional therapy does not necessarily

correspond with an immediate decrease in size of the treated lesion. The degree of necrosis seen on imaging is an important component in assessing early response to treatment [9].

The efficacy of transarterial chemoembolization and radiofrequency ablation has been evaluated by estimating the degree of necrosis on CT that correlated with the degree of necrosis on pathologic examination [20]. Furthermore, actual tumor necrosis within resected tumors corresponded to areas of nonenhancing tumor necrosis on post-radioembolization CT imaging [18]. In this small pilot study, Kaplan–Meier curves suggested correlation of tumor necrotic volume and subsequent survival when a >10% increase in tumor-necrosis volume was demonstrated on enhanced CT 1 month after radioembolization. A cut-off point of 10% increase in necrosis was chosen arbitrarily. In larger studies, different cut-off points might be evaluated for their predictive value. It is interesting that there was no statistically significant difference in survival times when comparing the group of patients with change of summated long-axis measurements <20% or ≥20% and those with change of whole tumor volume <20% versus ≥20% (p-value of log-rank test = 0.3198 and 0.4506, respectively). This further suggests shortcomings of RECIST criteria and supports the use of EASL assessment criteria focused on tumor necrosis when evaluating response to radioembolization [9]. However, to further evaluate the predictive value of volumetric change analysis, this technique should be applied in a large prospective clinical trial, as suggested by Sargent et al. [21].

It has been shown that necrosis measurements allowed significantly earlier detection of response relative to size criteria (World Health Organization and RECIST) because necrosis usually occurs shortly after treatment and before there is a decrease in lesion size. Miller et al. demonstrated necrosis criteria to be more reliable than size criteria for response evaluation in initial follow-up examinations after treatment [22]. The early identification of nonresponders allowed a prompt change to another therapy, such as chemoembolization [22]. Miller et al. demonstrated a median time to response of 29 days when evaluating tumor necrosis with enhanced CT. In the current study, the first follow-up imaging was obtained at 1 month to determine if results at this early time point might correlate with

subsequent survival. Functional MRI obtained at 1 month after radioembolization has been shown to predict response and correlate with survival [23]. However, it is not known if earlier or later imaging could demonstrate tumor response, which might be more predictive of treatment efficacy and survival outcomes.

In the case of chemoembolization, it has been shown that the extent of tumor necrosis, as demonstrated on CT 1 to 3 weeks after treatment, correlated with the extent of iodinized oil retention [24]. In our pilot study, we did not find a correlation between the delivered dose of ^{90}Y and subsequent tumor necrosis, tumor response, or survival. However, the delivered dose was based on hepatic lobar volume and percentage of tumor occupied by tumor, unlike the dose of chemotherapy delivered during chemoembolization.

Intraobserver reproducibility may be somewhat less than desirable when measuring viable enhancing tumor. One observer had a significant difference between the first and second set of measurements ($p = 0.004$). This occurred because the enhanced tumor portions are not different enough from attenuation values of the adjacent liver parenchyma, particularly in less-defined nonencapsulated and infiltrating tumors. Nevertheless, good interobserver reproducibility was demonstrated. Hunt et al. also demonstrated that contrast-enhanced CT or MRI obtained approximately 1 month after TACE and before transplantation was associated with a high error rate for the detection of viable tumor compared with histopathology [16]. Additional preprocessing may allow improved quantitation of the enhancing portion of the tumor. It is also true that the current standard of linear measurements of tumor diameters is limited when the tumor margins are ill defined, as with an infiltrating tumor.

This retrospective pilot study is limited by small sample size; variation in the patient population related to clinical presentation; tumor type, size, and number; and number of treatments. The sample size was too small to allow for stratification based on the factors that would affect survival outcomes. Determinants of therapeutic response are complicated and numerous and were not taken into account. Imaging-based tumor response does not by itself suggest clinically meaningful therapeutic efficacy unless there is also improvement in other clinical end points, such as survival or even quality of life. We therefore attempted initial assessment to determine if there was any correlation of these imaging end points with survival time for these patients. The analysis tool was applied retrospectively to CT studies for patients who had previously undergone their course of treatment. These analysis tools next need to be applied in a large, prospective, randomized clinical trial to evaluate how volumetric assessment of whole-tumor and tumor-necrosis volumes might predict survival outcomes. Once validated, volumetric analysis may then be used as an early surrogate marker of response in clinical trials and applied to clinical care decisions [21].

The current study design was not meant to evaluate accuracy. We previously demonstrated accuracy of the IWS approach using phantoms [8]. Others have compared imaging-based volumetric measurements of tumors, organs, and necrosis with resected tissue. In our medical center, few transplants are performed, and very few hepatic tumors are resected after local therapy. Therefore, we could not be certain that hypodense tumor tissue was necrotic or hypoperfused. However, a number of studies have demonstrated that "necrotic" tumor tissue on CT corresponded to actual tumor necrosis of subsequently resected tissue [18, 20].

A possible limitation of the semiautomated approach is that the operator must select the desired computer-generated boundaries. This can become a challenge at times, especially considering partial volume effects in the boundary areas (Fig. 8) [3]. Even when making linear manual whole-tumor measurements, the radiologist must determine what he or she thinks is the edge of the lesion. This may be difficult, particularly in the case of infiltrating ill-defined lesions, thus introducing operator-dependent error and interobserver/intraobserver variability, which decrease with semiautomated segmentation approaches. A benefit of this semiautomated segmentation, compared with a completely automated method, is that the operator has the ability to accept, decline, or modify the computer-generated outline. Such flexibility is often needed when

Fig. 8 (Left) Nonenhanced CT axial image through the abdomen with observer-set markers placed on a liver that directly abuts the stomach. (Right) Image output after IWS processing shows an uncertain border between liver and stomach, a problem likely attributed to partial volume averaging
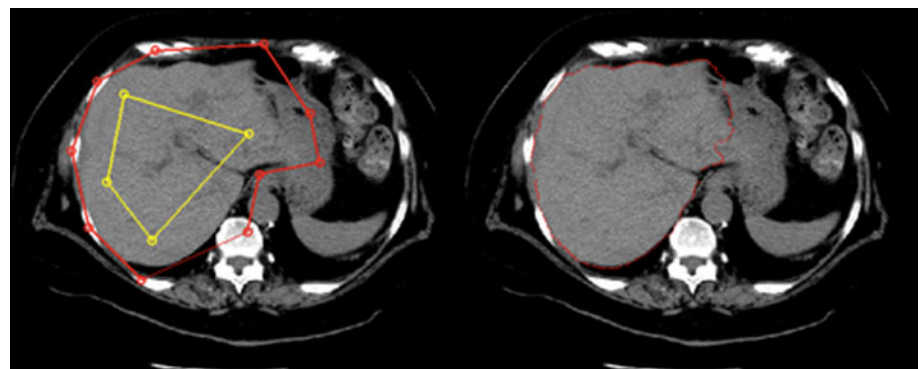
Fig. 9 (*Left*) Nonenhanced CT axial image through the abdomen with observer-set markers placed on a liver that splits on either side of the falciform ligament. (*Right*) Image output after IWS processing shows an overestimated area of the liver, including some of the fat-containing falciform ligament between the right and left lobe of the liver
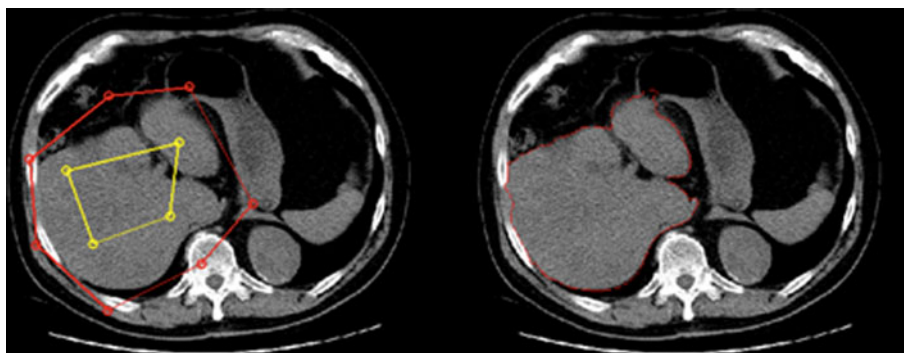
measuring volumes of the liver in close proximity to adjacent structures having similar Hounsfield units, such as the stomach or heart (Fig. 8).

A large number of segmentation approaches have been described for various applications [25–28]. In the approach presented, we did not use geometric modeling, such as spheres, to obtain segmentation maps, as described by Salvi et al. [28]. We did not use shape constraint to prevent the region growing from leakage into surrounding tissues of similar characteristics, as described by Zhoa et al. [27]. However, this approach may be helpful when the tumor is infiltrative or when the border is not easily discernable from adjacent parenchyma.

Another limitation occurs when trying to measure lobes of the liver separated by the fat-containing falciform ligament. This problem is currently approached by either bridging the split-liver areas (Fig. 9), or by measuring each segment separately. Lobar liver volumes were not addressed in this study. The measurement of segmental or lobar liver volumes using the proposed method can be performed. The volume of each lobe could be calculated separately by including only the lobe of interest in the outer boundary.

Several sources of variability affected reproducibility and accuracy in this study. The operators decided the manual placement of the inner and outer boundary markers. The best results were obtained when the inner boundary included a large and representative sample of liver parenchyma and the outer boundary excluded the liver parenchyma in its entirety. The operators could have also chosen to exclude nonhepatic tissue from the measurement. This is performed by placing an outer marker directly over the middle of unwanted tissue. Figure 2 demonstrates exclusion of the inferior vena cava by way of this technique.

In conclusion, good interobserver/intraobserver observer reproducibility of semiautomated volumetric assessment of liver, whole-tumor, and necrotic-tumor volumes is demonstrated. The technique is faster than traditional manual tracing methods. Although the described method of determining liver volume was tested on patients who were subsequently treated by radioembolization, the approach can be applied for planning liver surgery [29], resection [30], or liver transplantation [31]. This small pilot study may suggest that volumetric assessment of tumor necrosis could allow early prediction of subsequent survival outcomes. However, this must be further validated in larger prospective studies.

References

1. Salem R, Thurston KG (2006) Radioembolization with [90]Yttrium microspheres: a state-of-the-art brachytherapy treatment for primary and secondary liver malignancies. Part 1. Technical and methodologic considerations. J Vasc Interv Radiol 17:1251–1278
2. Nakayama Y, Li Q, Katsuragawa S et al (2006) Automated hepatic volumetry for living related liver transplantation at multisection CT. Radiology 240:743–748
3. Hermoye L, Laamari-Azial I, Cao Z et al (2005) Liver segmentation in living liver transplant donors: comparison of semiautomatic and manual methods. Radiology 234:171–178
4. Eisenhauer EA, Therasse P, Bogaerts J et al (2009) New response evaluation criteria in solid tumors: revised RECIST guideline (version 1.1). Eur J Cancer 45:228–247
5. Forner A, Ayuso C, Varela M et al (2009) Evaluation of tumor response after locoregional therapies in hepatocellular carcinoma: are response evaluation criteria in solid tumors reliable? Cancer 115:616–623
6. Kamel IR, Bluemke DA, Eng J et al (2006) The role of functional MR imaging in the assessment of tumor response after chemoembolization in patients with hepatocellular carcinoma. J Vasc Interv Radiol 17:505–512

7. Zhao B, Schwartz LH, Moskowitz CS et al (2006) Lung cancer: computerized quantification of tumor response—initial results. Radiology 241:892–898

8. Ray S, Hagge R, Gillen M et al (2008) Comparison of two-dimensional and three-dimensional iterative watershed segmentation methods in hepatic tumor volumetrics. Med Phys 35:5869–5881

9. Bruix J, Sherman M, Llovet JM et al (2001) Clinical management of hepatocellular carcinoma. Conclusions of the Barcelona 2000 EASL conference. European Association for the Study of the Liver. J Hepatol 35:421–430

10. Drever L, Robinson DM, McEwan A et al (2006) A local contrast based approach to threshold segmentation for PET target volume delineation. Med Phys 33:1583–1594

11. Fisher RA (1954) Statistical methods for research workers, 12th edn. Biological monographs and manuals. Hafner, New York, NY, 356 pp

12. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1(8476):307–310

13. Yohai VJ (1987) High breakdown point and high efficiency robust estimates for regression. Ann Stat 15:642–656

14. Cox DR (1972) Regression models and life tables. J R Stat Soc B 120:187–220

15. Kaplan EL, Meier P (1958) Nonparametric estimation from incomplete observations. J Am Stat Assoc 53:457–481

16. Hunt SJ, Yu W, Wientraub J et al (2009) Radiologic monitoring of hepatocellular carcinoma tumor viability after transhepatic arterial chemoembolization: estimating the accuracy of contrast enhanced cross sectional imaging with histopathologic correlation. J Vasc Interv Radiol 20:30–38

17. Keppke AL, Salem R, Reddy D et al (2007) Imaging of hepatocellular carcinoma after treatment with yttrium-90 microspheres. Am J Roentgenol 188:768–775

18. Riaz A, Kulik L, Lewandowski RJ et al (2008) Radiologic-pathologic correlation of hepatocellular carcinoma treated with internal radiation using yttrium-90 microspheres. Hepatology 49:1185–1193

19. Ibrahim SM, Nikolaidis P, Miller FH et al (2009) Radiologic findings following Y90 radioembolization for primary liver malignancies. Abdom Imaging 34:566–581

20. Pompili M, Mirante VG, Rondinara G et al (2005) Percutaneous ablation procedures in cirrhotic patients with hepatocellular carcinoma submitted to liver transplantation: assessment of efficacy at explant analysis and of safety for tumor recurrence. Liver Transpl 11:1117–1126

21. Sargent D, Rubinstein L, Schwartz L et al (2009) Validation of novel imaging methodologies for use as cancer clinical trials endpoints. Eur J Cancer 45:290–299

22. Miller FH, Keppke AL, Reddy D et al (2007) Response of liver metastases after treatment with Yttrium-90 microspheres: role of size, necrosis, and PET. Am J Roentgenol 188:776–783

23. Kamel IR, Reyes DK, Liapi E et al (2007) Functional MR imaging assessment of tumor after ^{90}Y microsphere treatment in patients with unresectable hepatocellular carcinoma. J Vasc Interv Radiol 18:49–56

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

25. Campadelli P, Casiraghi E, Esposito A (2009) Liver segmentation from computed tomography scans: a survey and a new algorithm. Artif Intell Med 45:185–196

26. McNitt-Gray MF, Bidaut LM, Armato SG et al (2009) Computed tomography assessment of response to therapy: tumor volume change measurement, truth data, and error. Transl Oncol 2:216–222

27. Zhao B, Schwartz LH, Jiang L et al (2006) Shape-constraint region growing for delineation of hepatic metastases on contrast-enhanced computed tomograph scans. Invest Radiol 41:753–762

28. Salvi E, Cantele F, Zampighi L et al (2008) JUST (Java User Segmentation Tool) for semi-automatic segmentation of tomographic maps. J Struct Biol 161:287–289

29. Tu R, Xia LP, Yu AL et al (2007) Assessment of hepatic functional reserve by cirrhosis grading and liver volume measurement using CT. World J Gastroenterol 13:3956–3961

30. Okamoto E, Yamanaka N, Oriyama T et al (1994) Prediction of the safe limits of hepatectomy by combined volumetric and functional measurements in patients with impaired hepatic function. Cancer Treat Res 69:293–299

31. Kawasaki S, Makuuchi M, Matsunami H et al (1993) Preoperative measurement of segmental liver volume of donors for living related liver transplantation. Hepatology 18:1115–1120