Interchangeability between real and three-dimensional simulated lung tumors in computed tomography: an interalgorithm volumetry study

Marthony Robins
Justin Solomon
Jocelyn Hoyer
Taylor Smith
Yuese Zheng
Lukas Ebner
Kingshuk Roy Choudhury
Ehsan Samei

Interchangeability between real and three-dimensional simulated lung tumors in computed tomography: an interalgorithm volumetry study

Marthony Robins,^{a,b,c,*} Justin Solomon,^{a,b,c} Jocelyn Hoyer,^{a,b,c} Taylor Smith,^{a,b,c} Yuese Zheng,^{a,c} Lukas Ebner,^{c,d} Kingshuk Roy Choudhury,^{a,c} and Ehsan Samei^{a,b,c}

^aCarl E. Ravin Advanced Imaging Laboratories, Durham, North Carolina, United States

^bDuke University, Medical Physics Graduate Program, Durham, North Carolina, United States

^cDuke University Medical Center, Department of Radiology, Durham, North Carolina, United States

^dUniversity of Bern, Department of Diagnostic, Interventional and Pediatric Radiology Inselspital, Bern, Switzerland

Abstract. Using hybrid datasets consisting of patient-derived computed tomography (CT) images with digitally inserted computational tumors, we establish volumetric interchangeability between real and computational lung tumors in CT. Pathologically-confirmed malignancies from 30 thoracic patient cases from the RIDER database were modeled. Tumors were either isolated or attached to lung structures. Patient images were acquired on one of two CT scanner models (Lightspeed 16 or VCT; GE Healthcare) using standard chest protocol. Real tumors were segmented and used to inform the size and shape of simulated tumors. Simulated tumors developed in Duke Lesion Tool (Duke University) were inserted using a validated image-domain insertion program. Four readers performed volume measurements using three commercial segmentation tools. We compared the volume estimation performance of segmentation tools between real tumors in actual patient CT images and corresponding simulated tumors virtually inserted into the same patient images (i.e., hybrid datasets). Comparisons involved (1) direct assessment of measured volumes and the standard deviation between simulated and real tumors across readers and tools, respectively, (2) multivariate analysis, involving segmentation tools, readers, tumor shape, and attachment, and (3) effect of local tumor environment on volume measurement. Volume comparison showed consistent trends (9% volumetric difference) between real and simulated tumors across all segmentation tools, readers, shapes, and attachments. Across all cases, readers, and segmentation tools, an intraclass correlation coefficient = 0.99 indicates that simulated tumors correlated strongly with real tumors ($p = 0.95$). In addition, the impact of the local tumor environment on tumor volume measurement was found to have a segmentation tool-related influence. Strong agreement between simulated tumors modeled in this study compared to their real counterparts suggests a high degree of similarity. This indicates that, volumetrically, simulated tumors embedded into patient CT data can serve as reasonable surrogates to real patient data. © 2018 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.JMI.5.3.035504]

Keywords: CT simulation; segmentation; lung tumor; quantitative; volume.

Paper 18092R received May 3, 2018; accepted for publication Aug. 27, 2018; published online Sep. 24, 2018.

1 Introduction

Quantitative computed tomography (CT) is gaining utility in clinical decision-making. Extraction of three-dimensional (3-D) tumor size characteristics has been shown to be a useful biomarker for disease diagnosis and treatment, and serves an equally important role in determining response to therapy.^{1,2} Quantitative assessments of tumor attributes can be impacted by the imaging system and image acquisition protocols.^{3,4} This impact is best investigated by clinical trials. However, these trials are often unfeasible and prohibitively expensive while, at the same time, ground truth is often unknown. An alternative to clinical trials is to use hybrid datasets, in which simulated tumors are digitally inserted into actual patient CT images. This approach is more practical, less expensive, and offers the advantage of known ground truth. Hybrid datasets can offer a viable alternative to answer a range of clinically relevant research questions,^{5–11} including eye tracking and decision-making

processes in radiologist's detection of lung nodules,⁸ task-specific image quality and detection of liver tumors,⁹ and validation of breast mass simulation algorithms.¹⁰

While simulated lesions have proven useful for many applications, their utility for quantification of tumor volume needs to be established and further verified against real lesions. Toward that goal, a prior study ascertained whether simulated tumors could serve as surrogates to real tumors in the context of a physical thoracic phantom.¹² Two image-based and one projection-based insertion methods^{13,14} simulated simple geometrical lesions, which were also physically implanted inside the thoracic phantom as the gold standard. Volumes of the inserted lesions, ascertained by a reader, were found to closely match the gold standard. While promising, however, this validation study did not offer the complexity of clinical thoracic backgrounds and lesions, nor did it incorporate the variability introduced by the readers and the segmentation algorithms.

The purpose of this study was to validate similarity between real and simulated lung tumors in the context of anthropomorphic

*Address all correspondence to: Marthony Robins, E-mail: marthony.robins@duke.edu

tumors inserted into clinical patient data, taking into consideration variability associated with readers and commercial segmentation algorithms. From the three insertion techniques used in the prior study,¹² only the one feasible for tumor insertion into reconstructed clinical datasets was deployed (technique B). Volumetric likeness was assessed by creating 3-D models of real tumors and digitally inserting them into the patient CT images containing the real tumors. Four independent readers then deployed three segmentation algorithms to measure tumor volumes. Similarities between real and simulated tumors were determined based on their absolute difference, multivariate analysis, and repeatability. The volumetric comparison of hybrid datasets was used to establish the interchangeability between real and simulated tumors.

2 Materials and Methods

2.1 Image Data

Patient image data were taken from the reference image database to evaluate therapy response (RIDER) dataset (Memorial Sloan Kettering Cancer Center).⁴ This dataset originally consisted of 32 pairs of lung CT examinations of patients with advanced lung cancer, acquired in a test–retest format. Patient images were acquired using a 16-detector row CT (Lightspeed 16; GE Healthcare, Milwaukee, Wisconsin) or 64 detector row CT (VCT; GE Healthcare). The scans were acquired using standard chest techniques given in Table 1.

Subjects were imaged during a single breath hold without the use of intravenous contrast agents. Images were reconstructed with 1.25-mm slice thickness (no overlaps or gaps) with filtered back projection, using a lung convolution kernel. Of the 32 cases available in the original database, 30 were included in this study (two cases were deemed unfit for modeling due to the inability to confidently discern the tumor). These 30 cases each contained at least one pathologically confirmed malignant tumor (non-small cell lung cancer). One tumor per subject was selected for modeling, most of which were primary lung tumors.¹⁵ Tumors were either isolated or had various degrees of attachment to the pleura, mediastinum, or vascular structures (see Fig. 1). In addition, some tumors were well-circumscribed while others had complex spiculated geometries. The approximate tumor volumes ranged from 0.764 to 136.4 cm³.

2.2 Tumor Segmentation

Tumors were segmented using an open-source segmentation software package (Seg3D, Center for Integrative Biomedical Computing, University of Utah, Salt Lake City, Utah). Thirty tumors were segmented in consultation with an experienced chest radiologist. The edges of each tumor were identified using a series of built-in tools and algorithms (e.g., threshold, connected components, dilate-erode, and Boolean functions).

Table 1 Thoracic CT protocol parameters used for GE Lightspeed 16 and VCT imaging systems.

CT system	kVp	mA	Exposure time (s)	Collimation (mm)	Pitch
Lightspeed 16	120	299–441	0.5–0.65	16 × 1.25	1.375
VCT	120	298–351	0.5–0.65	64 × 0.63	0.975

Fig. 1 Patient images from the RIDER dataset. (a)–(d) Tumors with various degrees of attachment to lung structures. Tumors were either (a) isolated, or attached to (b) the mediastinum, (c) vascular structures, or (d) the pleura.

Variability in segmentation was minimized by using these systematic tools, as opposed to manual contouring. Binary segmentation masks were then exported and used to create CT data volumes of interest (VOIs), from which tumor models were simulated. The tumor segmentation, modeling, and insertion scheme is shown in Fig. 2.

2.3 Tumor Modeling

The binary segmentation masks along with the VOIs were fed into a previously developed and validated tool (Duke Lesion Tool, Duke University) to generate computational tumor models corresponding to each real tumor. To enable mathematical tumor model development, Duke Lesion Tool is outfitted with an iterative fitting routine described by Solomon and Samei.¹⁶ Due to the complexity of many tumors, the goal was not to achieve an exact match between real tumors and their corresponding simulations but rather to develop simulated tumors that were reasonably similar in size and shape.

2.4 Creating Hybrid Datasets

Each simulated tumor was manually inserted into the image from which it was segmented and modeled. This insertion was done into the image domain in duplicate. These lesions were inserted at locations that were similar to the immediate environment of its corresponding real tumor. For example, if the real tumor was attached to the left chest wall, the corresponding tumor model duplicates were inserted such that they were attached to the left chest wall in a similar fashion. The simulated tumors were inserted via a previously validated 3-D image domain insertion Technique B¹² (see Fig. 3). This modified insertion process implemented a form of the alpha-blending technique to embed the lesion model into the patient CT data. First, the lesion model was voxelized at the resolution of the patient CT data. Then, upon identifying an appropriate VOI, a noise-only map was created for appropriate pixel value shifting with the voxelized lesion model. Using an alpha map of

Fig. 2 Flowchart depicting the multistep process for developing hybrid datasets. Real tumors are segmented from reconstructed CT data. Their size and shape are modeled, and virtually inserted into the original CT data.

Fig. 3 Image of a virtually inserted tumor with pleural attachment performed via image domain insertion.

the target patient CT data, anatomical structures in the target were removed before adding the lesion model. As a result, each hybrid patient case contained one real tumor and two virtually inserted tumors. The volumes of simulated tumors were approximately equivalent to those of real tumors. An experienced radiologist inspected all inserted tumors for similarity in tumor appearance and attachment type.

2.5 Volume Estimation

For each tumor (real and simulated), volumes were estimated by four independent readers in a fully crossed study approach. Each reader provided volume estimations for each tumor based on three commercial segmentation tools (iNtuition, TeraRecon, Inc., Foster City, California; Syngo.via, Siemens Healthcare, Germany; and IntelliSpace, Philips Healthcare, Andover, Massachusetts) (Fig. 4). This led to 1080 volume measurements {[30 real tumors + 60 simulated tumors] × 4 readers × 3

Table 2 Stratification of all real tumors in terms of shape and attachment. Some tumors had multiple types of attachment.

Shape		Attachment			
Well-					
Spiculated	circumscribed	Isolated	Pleura	Mediastinum	Vasculature
16	14	6	13	5	9

segmentation tools}. Table 2 provides an outline of the tumor shape and attachment classifications. The aforementioned software packages included semiautomated algorithms with the option for postsegmentation manual correction. Each segmentation tool provided readers with options to measure tumor size via single click (automated), region growing, or manual ROI. Readers used these basic algorithms supplied by the segmentation tools for volume estimation. All readers had basic prior experience with image segmentation ranging from minimal (<6 months) to moderate (~2 years). Regardless of their experience level, all readers were given instruction in operating these segmentation tools.

2.6 Statistical Methods

A series of statistical analyses were performed on the volume measurements to assess volumetric interchangeability between real and simulated tumors. These analyses included: (1) direct univariate comparison of real versus simulated tumor volumes in terms of overall trends and variability, (2) multivariate statistical modeling to assess how various factors (reader, segmentation tool, tumor shape/attachment, and real versus simulated) affected measured tumor volumes, and (3) assessment of how the local tumor environment affected segmentation precision.

Fig. 4 Each reader performed volume estimations using all three semiautomated commercial volume estimation tools: (1) iNtuition (TeraRecon, Inc., Foster City, California), (2) Syngo.via (Siemens Healthcare, Germany), and (3) IntelliSpace (Philips Healthcare, Andover, Massachusetts).

These results were anonymized with respect to segmentation tools and readers.

2.6.1 Volume comparison

For the 1080 readings in this study, a univariate analysis was done to directly compare the correlation of measured volumes between simulated and real tumors. Volume measurements were log transformed and a linear fit was used to estimate the intra-class correlation coefficient (ICC) between real and corresponding simulated tumors (as measured by a given reader with a given tool). In addition, variability was compared between real and simulated tumors by calculating the standard deviation (SD) of volume measurements across readers and tools for each case (30 real and 30 simulated tumors).

2.6.2 Multivariate analysis

A multivariate analysis using the restricted maximum likelihood method (REML) was applied to all tumors. To determine the degree of similarity between real and simulated tumors, a series of subanalyses were performed that incorporated conditions relevant to the study (i.e., segmentation tools, readers, tumor shape, and degree of attachment). To achieve this, the mixed effects model [Eq. (1)] was fit to data (nlme R computing package version 3.1-131).¹⁷ The model was of the form

$$\log(V_{ijkl}) = \mu + b_i + c_j + k_2 + \text{Exp}_i + \text{Seg}_l + \text{Sh}_j + \text{Att}_j + \text{Exp}_{ik_2} + \text{Seg}_{lk_2} + \text{Sh}_{jk_2} + \text{Att}_{jk_2} + \varepsilon_{ijkl}, \quad (1)$$

where V_{ijkl} is the tumor volume measured by the i 'th reader ($i = 1, \dots, 4$) for the j 'th case ($j = 1, \dots, 30$) of the k 'th type ($k_1 = \text{real}$ or $k_2 = \text{simulated}$) with the l 'th segmentation tool ($l = 1, \dots, 3$). The model approximates this measured volume in terms of a baseline tumor volume μ , which represents an isolated, well-circumscribed, real tumor, visualized using segmentation tool B, and read by the reader with least experience. Each term in Eq. (1) belongs to one of three effects categories (i.e., random effects, fixed effects, or interaction terms). Random effects, which assume that individual effects are subject-specific, were modeled for reader b_i , and case c_j . Both terms were assumed to have an independent zero-mean Gaussian distribution with SD σ_r and σ_c , respectively. Fixed effects refer to the population average of each condition while holding all others constant. Fixed effects for reader experience (moderate), tumor type (simulated), type of segmentation (tools A or B), shape (spiculated), and attachment (mediastinal, pleural, or vascular) were evaluated. Fixed effects terms were: simulated (k_2), experience (Exp), segmentation tool (Seg), shape (Sh), and attachment (Att). In addition, the model included interaction terms that represent the effect of each experimental condition on simulated tumors. For example, Exp_{ik_2} represents the interaction between the experience level of reader i and the simulated tumor type k_2 . Finally, ε_{ijkl} is the measurement error that was assumed to have a zero-mean Gaussian distribution with SD σ_e independent of random effects.

2.6.3 Location repeatability analysis

The local environment of the tumor can have an impact on volume estimation. In this study, identical copies of the same simulated tumor (m_1 and m_2) were inserted into two distinct locations in each patient case. This made it possible to investigate the

impact of the local environment on the measurement of tumor volumes. This was done by analyzing the precision of the volume measurements across the same tumors at different locations. The intratumor precision was determined by estimating the percent repeatability coefficient (PRC).¹⁸ The PRC for the i 'th reader, j 'th tumor, and l 'th segmentation tool is a function of the within mean square (WMS_{ijl}) and repeatability coefficient (RC_{ikl}) defined as

$$\text{WMS}_{ijl}^{\text{within-tumor}} = \sum_{m=1}^2 [\log(V_{ijlm}) - \log(\overline{V_{ijl}})]^2, \quad (2)$$

$$\hat{\sigma}_{ijl} = \sqrt{\text{WMS}_{ijl}^{\text{within-tumor}}}, \quad (3)$$

$$\text{RC}_{ikl} = 1.96\sqrt{2\hat{\sigma}_{ijl}^2} = 2.77\hat{\sigma}_{ijl}, \quad (4)$$

$$\text{PRC}_{ikl} = [\exp(\text{RC}_{ikl}) - 1] \times 100. \quad (5)$$

PRC represents the expected variation for 95% of tumor duplicates for all readers and tools. A smaller PRC implies that the insertion location had a smaller impact on volume measurement.

3 Results

3.1 Volume Comparison

Figure 5 shows the relative volume comparison between real and simulated tumors. One thousand eighty data points show the collective volume measures for four readers using three segmentation tools. Simulated tumor volume measurements are largely well-centered on the $x = y$ line, indicating strong agreement between simulated tumors and their corresponding real tumors. A linear regression of these data points shows that there is a 9% difference between real and simulated tumors, suggesting that simulated tumors closely mimic real tumor volumetry. ICC of 0.99 calculated across all cases, readers, and segmentation tools indicates that simulated tumors correlate strongly with real tumors.

The average measurements for each case (real and simulated) across four readers and three tools in Fig. 6 shows the observed trend with tumor measurement variability relative to the associated tumor size for real and simulated tumors. As tumor size increased, the variability in tumor measurement increased for both real and simulated tumors. As indicated by error bars, on average, real tumors had greater measurement variability, whereas simulated tumor measurement variability changed comparably with real tumor variability. Figure 7 explicitly shows this strong correlation (0.83), demonstrating that the variability observed in simulated tumors is proportional to the variability observed in the real tumors upon which the simulated tumors were based. The overall variability was 10% in simulated tumor measures relative to real tumors. This systematic difference in variability was consistent across all tumors of varying size, edge type, and attachment condition.

3.2 Multivariate Analysis

Table 3 lists the multivariate analysis results for the effect of shape, attachment, and segmentation tool on tumor volume

Fig. 5 Log-log plot shows the volume comparison between real and simulated tumors. Each point represents a single case (i.e., both real and simulated tumor volume measurements). For any given point, the x-axis indicated the volume of the real tumor measurement and the y-axis indicated the mean volume of the corresponding simulated tumor. One thousand eighty data points represent volume measures for four readers using three segmentation tools. There was a 9% difference between real and simulated tumors. ICC was determined to be 0.99.

Fig. 7 Log-log plot shows measurement variability (SD) between real and simulated tumors. Each point represents a single case of both real and simulated tumor variability measurements. For any given point, the x-axis indicates the log (SD) of the real tumor measurement and the y-axis indicates the log (SD) of the corresponding simulated tumor. There was 10% difference between real and simulated tumor variability with a correlation coefficient of 0.83.

Fig. 6 Plot provides size-based depiction of measurement variability between real and simulated tumors. Simulated tumor measurement variability changed proportionally with real tumors.

Table 3 Multivariate analysis results showing the relative relationship of each term with respect to the baseline. Baseline is the volume measurement of an isolated, well-circumscribed, real tumor visualized using segmentation tool B, and reader 1. Quantities were log-transformed before statistical analysis.

	Term	Difference in volume measurement (relative to baseline)	Std. error	t-value	p-value
Fixed terms	Baseline	10.45	0.38	27.78	—
	Tool A	−0.03	0.02	−1.51	0.13
	Tool C	−0.01	0.02	−0.36	0.72
	Shape	−0.07	0.28	−0.26	0.79
	Moderate experience	0.13	0.03	4.08	< 0.001
	Mediastinal attachment	0.19	0.43	0.44	0.66
	Pleural attachment	0.45	0.35	1.28	0.2
	Vascular attachment	−0.04	0.41	−0.11	0.91
	Simulated	0.03	0.53	0.06	0.95
Interaction terms	Moderate experience: simulated	−0.07	0.02	−3.52	< 0.001
	Segmentation tool A: simulated	−0.01	0.02	−0.23	0.82
	Segmentation tool C: simulated	−0.01	0.02	−0.6	0.55
	Shape: simulated	−0.01	0.4	−0.01	0.99
	Mediastinal attachment: simulated	−0.04	0.61	−0.07	0.94
	Pleural attachment: simulated	−0.07	0.5	−0.15	0.88
	Vascular attachment: simulated	−0.05	0.58	−0.09	0.93

measurement. It describes how each factor affected tumor volume, relative to the baseline condition (isolated, well-circumscribed, real tumor visualized using segmentation tool B and reader 1). There was no significant difference between volume measurements of real versus simulated tumors (p -value = 0.95), nor was there a significant effect from segmentation tools (p -values of 0.13 and 0.72 for tools A and C, respectively). In addition, there was no significant interaction effect between simulated tumors and tumor shape (p -value = 0.99). So, segmentation tools exhibited similar volume measurement response with simulated tumors (spiculated and well-circumscribed) as they did with real tumors. The experience level of readers affected measurement accuracy. Moderately experienced readers generated smaller volume measurements for simulated tumors (p -value = 0.03) than real. However, interreader variability ($\sigma_r = 0.03$), as well as intertool variability ($\sigma_t = 0.03$), were each relatively small compared to measurement error ($\sigma_e = 0.18$).

3.3 Location Repeatability Analysis

Intratumor variations between identical copies of simulated tumors (inserted at different locations) were compared by tool and reader in Table 4. Across all tools and readers, the average PRC was 7%. Table 4 shows that the range of PRC values was 4.1% and 1.7% across reader-based and tool-based measurement comparisons, respectively. This implies that PRC was

Table 4 Results for PRC between corresponding copies of simulated tumors digitally inserted into the same dataset at different locations. PRC proportionally indicates the impact of insertion location on tumor volume measurement.

Term	Tool A	Tool B	Tool C	All tools
Reader 1 (%)	5.6	5.3	7.0	6.0
Reader 2 (%)	9.5	9.6	9.7	9.6
Reader 3 (%)	7.6	4.0	9.3	7.0
Reader 4 (%)	6.5	5.1	5.0	5.5
All readers (%)	7.3	6.0	7.7	7.0

affected more by readers than tools. Tool B had the smallest overall PRC, indicating that it was least likely to be influenced by the local tumor environment.

4 Discussions

While a substantial portion of chest CT exams are for patient follow-up scans due to indeterminate nodules and oncology patient assessment, early detection and classification of lung tumors has been shown to reduce lung cancer mortality.^{19–21}

To this end, lung screening exams have been widely adopted by most major medical organizations including the American College of Radiology, Fleischner Society, American Cancer Society, and American Lung Association.^{22–25} In addition, official policies to advance lung CT screening have been approved by the American College of Chest Physicians, American Thoracic Society, Center for Medicare and Medicaid Services, and US Preventative Services Task Force.²⁶ These official endorsements, as well as increasing patient and physician interest in early detection of abnormalities, have driven higher demand for accurate tumor measurement and classification. To meet this demand, and to maximize the potential impact on patient outcome, tumor measurement and classification must be precisely performed and optimized. For this reason, Lung-RADSTM and the Fleischner Society have provided guidelines to direct physician decision-making for optimal diagnosis.^{24,27}

Tumor measurement invokes the quantitative analysis of CT images using segmentation tools. Such analysis can be limited by patient and technical considerations. In the context of clinical studies, testing every possible combination of scan parameters on every patient is an impossible approach to ascertaining ideal segmentation tool performance. Even if it were possible to gather the appropriate patient population to gain statistical and clinical significance, such studies would be challenging given the existing legal limitations concerning patient radiation exposure. Phantom studies that rely on segmentation of phantoms and physical test objects can remedy this limitation but cannot realistically represent *in vivo* conditions. Hybrid datasets that circumvent the challenges of both clinical and phantom studies could provide unparalleled advantages.

We previously demonstrated a first-order validation showing that simulated datasets can closely mimic real datasets in the context of an anthropomorphic phantom.¹² A comparison of techniques A and B relative to technique C was shown to be comparable to the previously developed insertion technique C. Volumetric similarity among techniques A, B, and C differed from the ground truth phantom measurements by <3%. In addition, lesion shape deformation assessment using 3-D Hausdorff distance measurements showed that simulated lesions (techniques A, B, and C) were 5% different from CT-derived phantom lesions. While these results were promising, the study was only performed using a phantom with overly simplified lesions and backgrounds. In addition, it only involved a single reader and a single segmentation algorithm. This study sought to address these limitations using real backgrounds, anthropomorphic tumors, and multiple segmentation algorithms and readers. From the prior insertion methods, the projection-based method (technique A) was not useable as the projection data from the clinical cases were not available. Technique C also could not accommodate potential overlapping anatomical backgrounds. Thus, the study was conducted using technique B, which offers maximum feasibility for broad applications.

Although there was no perfect match between real and simulated tumors, they realistically mimicked the characteristics of real tumors *in vivo* in terms of attachment and shape. Therefore, we sought to determine if simulated tumors were statistically similar to real tumors when replicating real tumor environment conditions. Simulated and real tumor volume measurements were highly similar across all tools and readers, with ICC of 0.99. This indicated that simulated and real tumor volumetry were comparable. In a prior phantom study, measured volumes of simulated

tumors were shown to be similar to those of synthetic tumors with <3% volumetric difference.²⁸ We observed a similar pattern of small volume difference between real and simulated lung tumors.

For each tool, simulated and real tumor volumes trended in the same way, suggesting volumetric interchangeability. Similarity between the real and simulated tumor volume estimation trends within each tool was evident, despite the heterogeneity of the 30 cases in terms of tumor size, shape, and type of attachment. Moreover, there were no statistically significant differences between real and simulated tumors when considering shape or attachment (Table 3). For example, shape-based comparison of real and simulated tumors showed that their volume measurements were decidedly similar (p -value = 0.79). When considered across all tools, comparison between real and simulated tumors showed that there was no statistically significant difference (p -value = 0.95). In Fig. 5, measurement of real and simulated tumor volumes across all volume estimation tools showed considerable overlap. Furthermore, for any given tool, results showed that the difference between real versus simulated tumor volume measurements was <9%. Essentially, volume estimation tool results showed strong similarity in volume measurement between real and simulated tumors. Similarly, comparison of variability in tumor measurement indicated a good correlation (0.83) between real and simulated tumors. Real tumor measurements, however, had a larger dynamic range across readers, generally reflecting greater variation relative to simulated tumors (10%) in terms of volume SD. Continued adjustments to the tumor simulation model are being pursued to minimize these differences.

One previous study has shown that increased lung complexity (e.g., the presence of scar tissue, vessels, and other lung structures) near tumors can have a distracting effect that impairs tumor detection and size perception.²⁹ To examine the influence of the local tumor environment on segmentation precision, we calculated PRCs to assess volume repeatability. For each case, we measured the volume of two identical copies of the simulated tumor inserted into different spaces in the patient image. As indicated in Table 4, the reader-based and tool-based PRCs of these measurements deviated by a maximum of $\pm 2.6\%$ and $\pm 1.0\%$, from the average PRC of 7%, respectively. These results indicate that volume measurements of simulated tumors were not strongly affected by the local tumor environment. The larger deviation among readers compared to tools suggest that readers were more susceptible to being impacted by the tumor's local environment.

A closer look at PRC values based on segmentation tools shows that tool B produced the smallest PRC, which suggested that the tumor background had the smallest effect on tumor measurement precision when using tool B. Conversely, tool C produced the largest PRC. This may be coupled with the fact that readers had little prior experience with this tool. Similarly, since each tool was used by the same readers, the ICC was used as a measure of reliability across tools. The paired readings agreed with a consistency of 0.99. Furthermore, the multivariate analysis indicated that variability across tools was relatively small ($\sigma_t = 0.03$). Tumor size contributed most to this measurement variability, with smaller-sized tumors producing larger volumetric fluctuations from baseline.

Many factors can influence the accuracy and efficiency of pulmonary tumor segmentation. Chen et al., Young et al., and Xie et al. have pointed out that volume measurement can be affected by a host of primary image acquisition factors,

including slice thickness and reconstruction kernel.^{30–32} Tumor measurement can also be affected secondarily by reader variability and segmentation algorithms.^{3,4,33} Since we did not acquire the source images for the RIDER dataset, we will only consider these potential secondary factors. First, the extent of manual editing of semiautomated volume estimations was not constrained. In all measurements, segmentation was initiated by a single click. Overall, readers preferred the automatic single click segmentation but reported that each segmentation tool failed to adequately characterize many tumors on single click. Some tools initially captured <50% of the tumor volume. Each tool required multiple manual adjustments after initial segmentation. This was followed by a manual correction in ~90% of cases. This could have contributed to the overall measurement variability. However, we permitted post-segmentation correction to optimize final volume estimates, particularly because of the large size of many tumors. As established in another study of patient cases segmented with semiautomated tools, manual correction substantially improved segmentation accuracy.³ In our study, simulated tumors were manually corrected in much the same way and in the same cases as real tumors. Second, the prior experience of readers was considered as a potential factor influencing measurement variability. In addition, a considerable number of tumors (>52%) had irregular contours. While readers had various levels of experience performing segmentation tasks, some readers had limited experience in operating some segmentation tools used in this study. Tool C was found to be the most intuitive and easy to manipulate for readers. Also, readers indicated that both tools B and C provided the best edge characterization via manual adjustment.

Tumor border discernment proved challenging for readers. Particularly, real tumors had greater boundary ambiguity due to a lack of pixel intensity distinction. When tumors were attached to lung structures, this made the task of segmentation especially difficult and subjective. Readers had to rely on manual intermediation methods such as adjusting window and level settings to assist them in distinguishing between tumor and normal lung tissue. In such cases, the automatic tool functions failed to adequately characterize tumor boundaries autonomously. Conversely, some isolated tumors (i.e., tumors without any attachments) were segmented by single-clicks only.

Differences in volume estimates between segmentation tools could be related to their unique algorithms. Since the segmentation tool internal functions were not disclosed, we treated them as black boxes. Interreader variability was trivial for both simulated and real tumors ($\sigma_r = 0.03$) and was relatively small compared to the measurement error ($\sigma_e = 0.18$). Moreover, the volumetry results for simulated tumors in this study were congruent with Wormanns et al.'s study, which demonstrated that interreader variability is almost negligible compared to the variability caused by other factors.³⁴

The current study shares inherent limitations related to noise and resolution. Real tumors rendered in the CT data are not ideal for simulation purposes. Since the simulated tumors were modeled based on the real tumors, they were inherently detail-limited. To preserve as much detail as possible, we used a continuous function to fit the tumor models to the real tumors. During this process, the simulated tumors can potentially be subject to over smoothing and incur additional loss of detail. This could have contributed to reduced variability in tumor

edge depiction. In this regard, an alternative method to capture greater tumor detail would be micro-CT. The superior resolution of micro-CT can guarantee better edge detail for improved edge characterization and lesion modeling. Also, segmentation-specific shape assessment was not possible since the commercial algorithms used did not permit the output of segmentation masks. Although exactly mimicking the real tumors was not our goal in this study, a challenge with mimicking real tumors occurs when there is significant tumor attachment to surrounding lung structures. This challenge is attributed to the fact that the modeling process does not extend beyond the tumor to its macroenvironment. For this purpose, development of a tumor growth model to account for the dendritic rendering of scar tissue resulting from invasive tumors would be beneficial. Even so, quantitative assessments of tumor attributes are often impacted by technical limitations related to the imaging system and image acquisition protocols. Hybrid data can offer realistic image quality for many novel studies, from the optimization of protocols, to the assessment of segmentation algorithms across different imaging systems.

5 Conclusion

This study serves as a validation of statistical interchangeability between real and simulated tumors, using a tumor modeling and virtual insertion approach. The results show that simulated tumors embedded in clinical patient images can quantitatively (i.e., volume and shape) be used as reasonable surrogates for patient CT datasets. Such hybrid datasets allow an unprecedented opportunity to develop a wide variety of cases for virtual clinical trials to test the impact of imaging conditions on lesion quantitation, maximize study sample size, circumvent the financial, ethical, and safety limitations of repeatedly acquiring patient CT data, and minimize the quantitative uncertainty resulting from not knowing tumor ground truth. Moreover, this approach has great potential to precede and augment preclinical investigations of new and developing CT technologies.

Disclosures

The authors have no relevant financial interests in the manuscript and no other potential conflicts of interest to disclose.

Acknowledgments

This work was supported by the National Institutes of Health (NIH) through the Quantitative Imaging Biomarkers Alliance (QIBA) of the Radiological Society of North America (RSNA) Grant No. HHSN268201500021C. Also, the authors would like to acknowledge Mrs. Susan Whitney for assistance with access to the segmentation software.

References

1. P. Therasse et al., "New guidelines to evaluate the response to treatment in solid tumors," *J. Natl. Cancer Inst.* **92**(3), 205–216 (2000).
2. H. A. Gietema et al., "Pulmonary nodules: interscan variability of semi-automated volume measurements with multisection CT—influence of inspiration level, nodule size, and segmentation performance," *Radiology* **245**(3), 888–894 (2007).
3. B. de Hoop et al., "A comparison of six software packages for evaluation of solid lung nodules using semi-automated volumetry: what is the minimum increase in size to detect growth in repeated CT examinations," *Eur. Radiol.* **19**(4), 800–808 (2009).
4. B. Zhao et al., "Evaluating variability in tumor measurements from same-day repeat CT scans of patients with non-small cell lung cancer," *Radiology* **252**(1), 263–272 (2009).

5. B. Chen et al., "Lesion insertion in projection domain for CT image quality assessment," *Proc. SPIE* **9412**, 94121R (2015).
6. X. Li et al., "Three-dimensional simulation of lung nodules for paediatric multidetector array CT," *Br. J. Radiol.* **82**(977), 401–411 (2009).
7. E. Samei, M. Flynn, and W. Eyler, "Simulation of subtle lung nodules in projection chest radiography," *Radiology* **202**(1), 117–124 (1997).
8. G. D. Rubin et al., "Characterizing search, recognition, and decision in the detection of lung nodules on CT scans: elucidation with eye tracking," *Radiology* **274**(1), 276–286 (2015).
9. J. Solomon et al., "Diagnostic performance of an advanced modeled iterative reconstruction algorithm for low-contrast detectability with a third-generation dual-source multidetector CT scanner: potential for radiation dose reduction in a multireader study," *Radiology* **275**(3), 735–745 (2015).
10. L. Sisternes et al., "A computational model to generate simulated three-dimensional breast masses," *Med. Phys.* **42**(2), 1098–1118 (2015).
11. M. Robins et al., "Inter-algorithm lesion volumetry comparison of real and 3D simulated lung lesions in CT," *Proc. SPIE* **10132**, 101321S (2017).
12. M. Robins et al., "Techniques for virtual lung nodule insertion: volumetric and morphometric comparison of projection-based and image-based methods for quantitative CT," *Phys. Med. Biol.* **62**(18), 7280–7299 (2017).
13. A. Pezeshk et al., "Seamless insertion of real pulmonary nodules in chest CT images," *Proc. SPIE* **9035**, 90351K (2014).
14. A. Pezeshk et al., "Seamless insertion of pulmonary nodules in chest CT images," *IEEE Trans. Biomed. Eng.* **62**(12), 2812–2827 (2015).
15. A. J. Buckler et al., "Inter-method performance study of tumor volumetry assessment on computed tomography test-retest data," *Acad. Radiol.* **22**(11), 1393–1408 (2015).
16. J. Solomon and E. Samei, "A generic framework to simulate realistic lung, liver and renal pathologies in CT imaging," *Phys. Med. Biol.* **59**(21), 6637–6657 (2014).
17. J. Pinheiro et al., "Linear and nonlinear mixed effects models," R package version 3.57, pp. 1–89 (2007).
18. H. X. Barnhart and D. P. Barboriak, "Applications of the repeatability of quantitative imaging biomarkers: a review of statistical analysis of repeat data sets," *Transl. Oncol.* **2**(4), 231–235 (2009).
19. L. Humphrey et al., "Screening for lung cancer: systematic review to update the US Preventive Services Task Force recommendation," Report No.: 13-05188-EF-1, Agency for Healthcare Research and Quality, Rockville (2013).
20. Z. Saghir et al., "CT screening for lung cancer brings forward early disease. The randomised Danish Lung Cancer Screening Trial: status after five annual screening rounds with low-dose CT," *Thorax* **67**(4), 296–301 (2012).
21. R. Siegel et al., "Cancer statistics, 2014," *CA Cancer J. Clin.* **64**(1), 9–29 (2014).
22. R. Wender et al., "American cancer society lung cancer screening guidelines," *CA Cancer J. Clin.* **63**(2), 106–117 (2013).
23. N. L. S. T. R. Team, "Results of initial low-dose computed tomographic screening for lung cancer," *N. Engl. J. Med.* **2013**(368), 1980–1991 (2013).
24. H. MacMahon et al., "Guidelines for management of incidental pulmonary nodules detected on CT images: from the Fleischner Society 2017," *Radiology* **284**(1), 228–243 (2017).
25. "ACR acknowledges ACS lung cancer screening guidelines: cautions standards needed to support widespread screening 2013," <https://www.acr.org/About-Us/Media-Center/Press-Releases/2013-Press-Releases/20130206ACR-Acknowledges-ACS-Lung-Cancer-Screening-Guidelines> (accessed 6 February 2013).
26. R. S. Wiener et al., "An official American Thoracic Society/American College of Chest Physicians policy statement: implementation of low-dose computed tomography lung cancer screening programs in clinical practice," *Am. J. Respir. Crit. Care Med.* **192**(7), 881–891 (2015).
27. "Lung-RADS Version 1.0 Assessment Categories 2014," <https://www.acr.org/~media/ACR/Documents/PDF/QualitySafety/Resources/Lung-RADS/AssessmentCategories.pdf> (accessed 28 April 2014).
28. M. Robins et al., "Development and comparison of projection and image space 3D nodule insertion techniques," *Proc. SPIE* **9783**, 97835X (2016).
29. L. Ebner et al., "Variations in the functional visual field for detection of lung nodules on chest computed tomography: impact of nodule size, distance, and local lung complexity," *Med. Phys.* **44**(7), 3483–3490 (2017).
30. B. Chen et al., "Volumetric quantification of lung nodules in CT with iterative reconstruction (ASiR and MBiR)," *Med. Phys.* **40**(11), 111902 (2013).
31. S. Young et al., "Variability in CT lung-nodule volumetry: effects of dose reduction and reconstruction methods," *Med. Phys.* **42**(5), 2679–2689 (2015).
32. X. Xie et al., "Sensitivity and accuracy of volumetry of pulmonary nodules on low-dose 16-and 64-row multi-detector CT: an anthropomorphic phantom study," *Eur. Radiol.* **23**(1), 139–147 (2013).
33. D. Han, M. A. Heuvelmans, and M. Oudkerk, "Volume versus diameter assessment of small pulmonary nodules in CT lung cancer screening," *Transl. Lung Cancer Res.* **6**(1), 52–61 (2017).
34. D. Wormanns et al., "Volumetric measurements of pulmonary nodules at multi-row detector CT: in vivo reproducibility," *Eur. Radiol.* **14**(1), 86–92 (2004).

Biographies for the authors are not available.