Evaluation of 1D, 2D and 3D nodule size estimation by radiologists for spherical and non-spherical nodules through CT thoracic phantom imaging

Nicholas Petrick[*][*][a], Hyun J. Grace Kim[b], David Clunie[c], Kristin Borradaile[c],
Robert Ford[d], Rongping Zeng[a], Marios A. Gavrielides[a], Michael F. McNitt-Gray[b],
Charles Fenimore[e], Z. Q. John Lu[e], Binsheng Zhao[f] Andrew J. Buckler[g]

[a]Center for Devices and Radiological Health, U.S. Food and Drug Administration, Silver Spring, MD
[b]David Geffen School of Medicine at UCLA, Los Angeles, CA
[c]CoreLab Partners, Inc, Princeton, NJ
[d]Princeton Radiology, Princeton, NJ
[e]National Institute of Standards and Technology, Gaithersburg, MD
[f]Department of Radiology, Columbia University Medical Center, New York, NY
[g]Buckler Biomedical LLC, Wenham, MA, USA

ABSTRACT

The purpose of this work was to estimate bias in measuring the size of spherical and non-spherical lesions by radiologists using three sizing techniques under a variety of simulated lesion and reconstruction slice thickness conditions. We designed a reader study in which six radiologists estimated the size of 10 synthetic nodules of various sizes, shapes and densities embedded within a realistic anthropomorphic thorax phantom from CT scan data. In this manuscript we report preliminary results for the first four readers (Readers 1-4). Two repeat CT scans of the phantom containing each nodule were acquired using a Philips 16-slice scanner at a 0.8 and 5 mm slice thickness. The readers measured the sizes of all nodules for each of the 40 resulting scans (10 nodules $\times$ 2 slice thickness $\times$ 2 repeat scans) using three sizing techniques (1D longest in-slice dimension; 2D area from longest in-slice dimension and corresponding longest perpendicular dimension; 3D semi-automated volume) in each of 2 reading sessions. The normalized size was estimated for each sizing method and an inter-comparison of bias among methods was performed. The overall relative biases (standard deviation) of the 1D, 2D and 3D methods for the four readers subset (Readers 1-4) were -13.4 (20.3), -15.3 (28.4) and 4.8 (21.2) percentage points, respectively. The relative biases for the 3D volume sizing method was statistically lower than either the 1D or 2D method (p<0.001 for 1D vs. 3D and 2D vs. 3D).

Keywords: Size Estimation, Lung Nodules, CT Imaging, Phantom Study, Reader Study, Reader Variability

1. INTRODUCTION

Computer tomography (CT) has made great advances over the past few decades enabling these newer CT scanners to acquire high speed, high spatial resolution image data. These advances have facilitated new potential applications for CT including lung cancer screening. Results reported in the literature indicate that CT may be an effective tool for early detection of lung cancer[1,2] with the National Lung Cancer Screening trial reporting that patients undergoing CT screening showed 20.3% fewer deaths compared with patient having planar x-ray imaging[1]. Another important use of CT imaging is monitoring the progression of lung cancer for patient undergoing therapy whether in drug trials or clinical practice[3]. While it is recognized that quantitative imaging (and in particular quantitative CT) will play an important role, significant efforts are still needed to identify appropriate imaging and measurement techniques required to maximize quantitative utility.

---

[*]nicholas.petrick@fda.hhs.gov, WO62-4118, 10903 New Hampshire Avenue, Silver Spring, MD 20993

As a way to systematically address limitations in quantitative imaging, the Radiological Society of North America (RSNA) formed the Quantitative Imaging Biomarker Alliance (QIBA) to investigate the role of quantitative imaging methods in CT, MRI and PET as potential biomarkers in evaluating disease and response to treatment[4]. The alliance formed technical committees of representatives from the instrumentation manufacturers, software developers, imaging professionals in the pharmaceutical industry, radiologists from the imaging contract research organizations (CROs), officers in regulatory agencies, governmental research organizations, imaging scientists, and professional imaging society representatives. One such technical committee is the volumetric CT (vCT) committee, which has taken on the task of developing quantitative tumor size measurement methods for CT. The current standard for monitoring solid nodules is known as RECIST[5,6]. RECIST is a latent standard for assessing diseases progression (or regression) in patients with lung cancer[5]. However, the RECIST measurement standard is limited to the longest, in-plane diameter of a tumor as a proxy for tumor size and is thought to be potentially problematic because tumors do not always expand or contract uniformly[7]. A solution to the limitations of RECIST may be the use of volume measurement tools[8], however questions have been raised about whether volumetric image analysis will add value or only increase the costs of patient care and the complexity of running clinical trials[3]. The QIBA vCT committee is working to develop groundwork data to help answer these questions. As part of these efforts, the QIBA vCT group is undertaking efforts to better quantify the accuracy and precision of nodule size measurements made by clinicians reviewing CT scans. The goal of this particular effort, known as Part 1A of the QIBA vCT Roadmap[9], is to estimate the bias and variance for radiologists estimating the 1D, 2D and volumetric size of synthetic nodules from CT scans of an anthropomorphic thorax phantom containing nodules embedded within the lung region of the phantom.

2. MATERIALS AND METHODS

2.1 Database Description

The study uses CT data collected from an anthropomorphic thorax phantom containing synthetic nodules of varying shapes and sizes acquired as part of a separate FDA research project[10,11]. The anthropomorphic thorax phantom (Kyotokagaku Incorporated, Tokyo, Japan) is shown in Fig. 1. It's designed to mimic the x-ray attenuation properties of the various tissues and structures within the thorax and includes a vascular inert to further mimic the complex vascular structure found within the lung region[10].
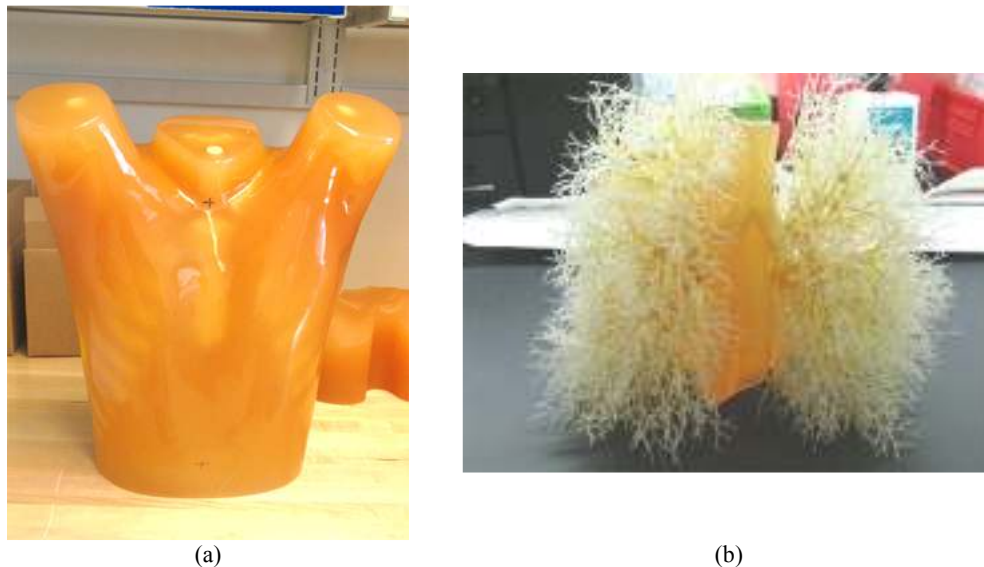

<div align="center">(a)   (b)</div>

Figure 1. Photo of the (a) anthropomorphic thorax phantom along with its (b) vascular insert. Synthetic phantom nodules were attached to the vascular insert and the phantom was scanned using a Philips 16-row Mx8000 IDT CT scanner (Philips Healthcare, Andover, MA).

A total of 10 nodules (five different shape/size combinations at two different CT densities) were inserted within the thorax phantom and imaged. Table 1 summarizes the characteristics of these nodules while Fig. 2 depicts the various nodules shapes evaluated by the readers. The thorax phantom was imaged using a Philips Mx8000 IDT 16-row scanner (Philips Healthcare, Andover, MA) with scan acquisition parameters summarized in Table 2. A total of 40 CT datasets (10 nodules $\times$ 2 slice thicknesses $\times$ 2 repeat exposures) were evaluated by the reviewing radiologists during the reader study.

Table 1. Table summarizing the characteristics of the synthetic nodules measured in this study. A total of 10 different nodules were evaluated include five different nodules shape/size combinations at two different densities. Equivalent diameter is defined as the diameter of a sphere with the same volume as that of the irregular shaped nodule.

Shape	CT Densities	Equivalent Diameter
Spherical	-10 HU, +100 HU	10 mm
Spherical	-10 HU, +100 HU	20 mm
Elliptical	-10 HU, +100 HU	20 mm
Lobulated	-10 HU, +100 HU	10 mm
Spiculated	-10 HU, +100 HU	10 mm

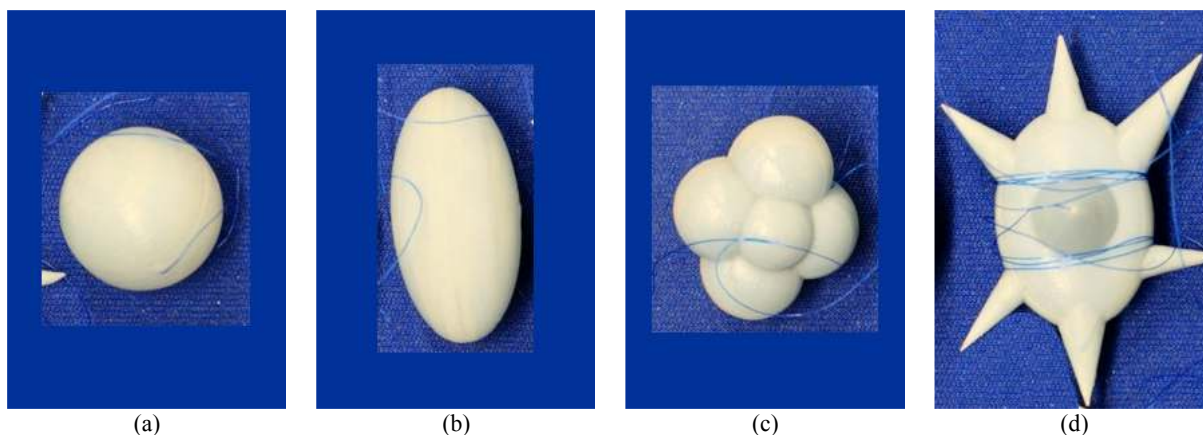(a)   (b)   (c)   (d)

Figure 2. Photos of the four different synthetic nodule shapes evaluated by the clinicians participating in the study. The four basic shapes are (a) spherical, (b) elliptical. (c) lobulated, and (d) spiculated.

• Table 2 Table summarizing the scan acquisition parameters used to acquire the CT image data of the thorax phantom.

Acquisition Parameter	Value
Scanner	Philips 16-slice MxIDT 8000
Tube Voltage	120 kVp
Exposure	100 mAs/slice
Pitch	1.2
Reconstruction Kernel	Detail
Slice Thickness	0.8 mm (16x0.75 mm collimation) 5.0 mm (16x1.5 mm collimation)
Repeat Exposures	2 scans for each nodule

2.2 Reading Protocol

Radiologists familiar with evaluating lesion response in drug trials participated in this reader study. They measured the size of a single nodule in 40 CT datasets using three different measurement techniques in each of two reading sessions. Each reading session was separated by at least 3 weeks. The sizing methods were: (1) a manual 1D uni-dimensional measurement using electronic calipers, (2) a manual bi-dimensional measurement using electronic calipers and (3) a 3D volumetric measurement using a semi-automated 3D volumetric tool.

The 1D technique measured the largest in-slice diameter for the lesion. This is loosely based on the RECIST criteria[5,6] as far as the measurement technique, however, it excludes summing over multiple nodules and any associated recommendations for assessing change. The 2D size measurement was performed in a similar fashion, but included a second step in which the largest perpendicular diameter within the same slice was also measured. These two linear measurements were then multiplied to provide an area-based size estimate. The 2D measurement is loosely based on the WHO criteria[12] but again limited to only the single lesion size measurement method. The 3D volumetric measurement was made using a proprietary semi-automated tool. The 3D measurement process required the reader to (1) define seed strokes, (2) apply the segmentation tool, (3) evaluate the quality of the segmentation, (4) refine (add/subtract) seed strokes and reapply the segmentation tool as necessary, and (5) repeat steps 3-4 until satisfied with the 3D segmentation of the nodule. The software then provided an estimate of nodule volume, which was recorded for the case. The semi-automated volume software also estimated 1D and 2D sizes using the results of the 3D segmentation but this data was not evaluated as part of this preliminary study.

All reading sessions took place at a CRO central facility (CoreLab Partners Inc., Princeton, NJ) using a proprietary software application and consumer color LCD monitors calibrated to the DICOM Grayscale Standard Display Function. Readers, cases, and measurement techniques were randomized to reduce biases in the study. The size measurements were made using a lung window/level display setting of 1200HU (window) and -600HU (level). The readers did not subsequently adjust the window width and level. All measurement techniques were performed on a single nodule within each of the 40 CT datasets and each measurement technique was independently applied (i.e., the manual 2D technique involved a separate estimate of the longest diameter instead of relying on the manual 1D technique estimate).

2.3 Analysis

We report and compare the relative bias and relative standard deviation for each measurement technique for four of the six readers (Readers 1-4) in this preliminary analysis. The use of a normalized metrics facilitates the comparison among the sizing techniques. Some type of normalization or conversion is necessary because each of the size measures is in a different dimensional space. The relative size, relative bias and relative standard deviation that we used in this analysis are defined as:

$$Size_{Rel} = \frac{Size_{Est} - Size_{True}}{Size_{True}} \bullet 100\% \quad \text{and} \qquad (1)$$

$$Bias_{Rel} = \text{Bias}(Size_{Rel}) \qquad (2)$$

$$Std_{Rel} = \frac{Std_{Est}}{Size_{True}} \bullet 100\% \qquad (3)$$

where $Size_{Est}$ and $Size_{True}$ is the estimated and true size, respectively. Std_{Est} is the estimated standard deviation for the size measuremen. We applied a mixed-effects linear regression model to the relative bias with nodule shape/size, nodule density and slice thickness as fixed effects and readers as a random effect. The regression model was used as the basis for statistical comparisons among the sizing methods.

3. RESULTS

The overall relative biases (standard deviation) of the 1D, 2D and 3D methods are given in Table 3. Fig. 3 shows box plots for the reader sizing results. The plots shown in Fig. 3 are averaged across all nodules, slice thicknesses and readers. We also evaluated the statistical difference in relative bias between pairs of the sizing methods (i.e., 1D vs. 2D, 1D vs. 3D and 2D vs. 3D) based on the regression model discussed above. This evaluation showed that the bias for the 3D method is statistically significantly smaller than the other two sizing method (p=0.072, p<0.001, and p<0.001, respectively).

• Table 3 The overall relative biases (standard deviation) of the 1D, 2D and 3D sizing methods averaged across all nodules, slice thicknesses and readers.

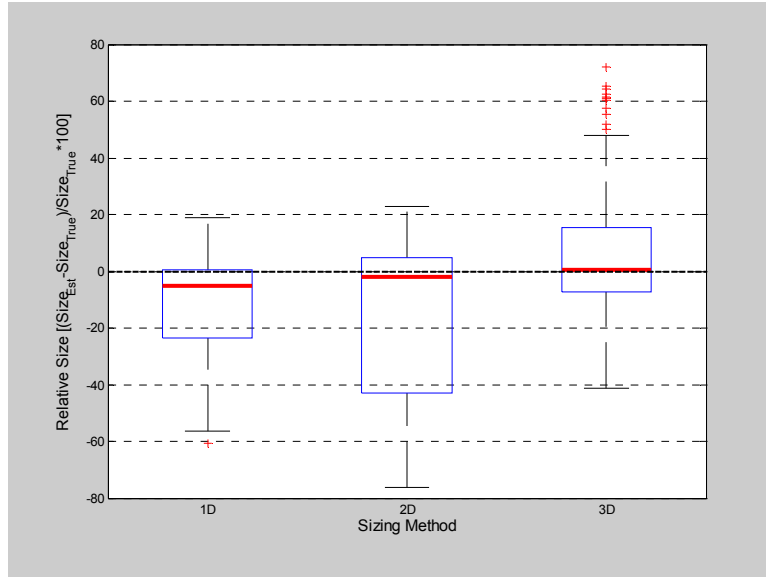| Sizing Methods	Relative Bias	Relative Standard Deviation
1D	-13.4%	20.3%
2D	-15.3%	28.4%
3D	4.8%	21.2%

Figure 3. Box plots of the distribution of reader sizing results for the 1D, 2D and 3D sizing methods. The distributions shown in Fig. 3 are for the combination of all readers sizing all nodules. The middle line in a box plot is the sample median; The tops and bottoms of each box correspond to the 25th and 75th percentiles, respectively.

4. DISCUSSION

The advantage of this type of phantom imaging study is that the true sizes of the nodules are known. This allows for both the bias and variance in reader sizing to be evaluated. The results from our preliminary analysis of four readers indicates that the 3D volume provides a low bias estimate of nodule volume while the 1D and 2D sizing techniques underestimate the longest diameter and largest cross-sectional area, respectively, for this group of nodules. The variability results show that precision of the size measurements was similar between the 1D and 3D measures with the 2D method having a larger relative standard deviation. One also notices a larger fraction of outliers in the box plots for the 3D size metric. This may result from a lack of flexibility in the semi-automated tool, which did not allow readers to manually adjust the segmented boundaries to their satisfactions. Evaluation of individual reader segmentations will be undertaken to provide additional insight into this observation.

A full analysis from all six total readers participating in the study has been undertaken and extends the preliminary analysis reported here by also analyzing data stratified by nodule characteristics and CT acquisition parameters. We are also developing additional regression and ANOVA models to understand how different nodule and acquisition parameters impact the estimability of lesion size. Finally, we are extending this basic study design to include clinical cases containing actual lung nodules to further investigate how biological factors impact the ability of readers to size lesions and, more importantly, track lesion size changes over time.

5. ACKNOWLEDGEMENTS

REFERENCES

[1]. "Lung cancer trial results show mortality benefit with low-dose CT: Twenty percent fewer lung cancer deaths seen among those who were screened with low-dose spiral CT than with chest X-ray". http://www.cancer.gov/newscenter/pressreleases/NLSTresultsRelease (2010).

[2]. International Early Lung Cancer Action Program, I., Henschke, C.I., Yankelevitz, D.F., et al., "Survival of patients with stage I lung cancer detected on CT screening.[see comment]". N Engl J Med, **355**(17), 1763-71 (2006).

[3]. Mozley, P.D., Schwartz, L.H., Bendtsen, C., et al., "Change in lung tumor volume as a biomarker of treatment response: a critical review of the evidence". Ann Oncol, **21**(9), 1751-5 (2010).

[4]. Buckler, A.J., Schwartz, L.H., Petrick, N., et al., "Data sets for the qualification of volumetric CT as a quantitative imaging biomarker in lung cancer". Opt. Express, **18**(14), 15267-15282 (2010).

[5]. Eisenhauer, E.A., Therasse, P., Bogaerts, J., et al., "New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1)". Eur J Cancer, **45**(2), 228-47 (2009).

[6]. Therasse, P., Arbuck, S.G., Eisenhauer, E.A., et al., "New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada". J Natl Cancer Inst, **92**(3), 205-16 (2000).

[7]. Buckler, A.J., Mulshine, J.L., Gottlieb, R., et al., "The use of volumetric CT as an imaging biomarker in lung cancer". Acad Radiol, **17**(1), 100-6 (2010).

[8]. Gavrielides, M.A., Kinnard, L.M., Myers, K.J., et al., "Noncalcified lung nodules: volumetric assessment with thoracic CT". Radiology, **251**(1), 26-37 (2009).

[9]. Buckler, A.J., Mozley, P.D., Schwartz, L., et al., "Volumetric CT in Lung Cancer: An Example for the Qualification of Imaging as a Biomarker". Acad Radiol, **17**(1), 107-115 (2010).

[10]. Gavrielides, M.A., Kinnard, L.M., Myers, K.J., et al., " A resource for the development of methodologies for lung nodule size estimation: database of thoracic CT scans of an anthropomorphic phantom". Optics Express, **18**(4), 15244-15255 (2010).

[11]. Gavrielides, M.A., Kinnard, L.M., Myers, K.J., et al., "FDA phantom CT database: a resource for the assessment of lung nodule size estimation methodologies and software development". SPIE Medical Imaging 2010: Computer-Aided Diagnosis, 7624, 762417-8 (2010).

[12]. [World Health Organization.WHO Handbook for Reporting Results of Cancer Treatment], Offset Publication No. 48 Geneva, Switzerland, (1979).