# A Mathematical Simulation to Assess Variability in Lung Nodule Size Measurement Associated With Nodule-Slice Position

Krishna Juluru • Noor Al Khori • Sha He • Amy Kuceyeski • John Eng

Published online: 20 December 2014 © Society for Imaging Informatics in Medicine 2014

Abstract The purpose of this study is to assess the variance and error in nodule diameter measurement associated with variations in nodule-slice position in cross-sectional imaging. A computer program utilizing a standard geometric model was used to simulate theoretical slices through a perfectly spherical nodule of known size, position, and density within a background of "lung" of known fixed density. Assuming a threshold density, partial volume effect of a voxel was simulated using published slice and pixel sensitivity profiles. At a given slice thickness and nodule size, 100 scans were simulated differing only in scan start position, then repeated for multiple node sizes at three simulated slice thicknesses. Diameter was measured using a standard, automated algorithm. The frequency of measured diameters was tabulated; average errors and standard deviations (SD) were calculated. For a representative 5-mm nodule, average measurement error ranged from +10 to -23 % and SD ranged from 0.07 to 0.99 mm at slice thicknesses of 0.75 to 5 mm, respectively. At fixed slice thickness, average error and SD decreased from peak values as nodule size increased. At fixed nodule size, SD increased as slice thickness increased. Average error exhibited dependence on both slice thickness and threshold. Variance and error in nodule diameter measurement associated with nodule-slice position

K. Juluru (⊠) • N. Al Khori • A. Kuceyeski Weill Cornell Medical College, 525 E. 68th St., F-056, New York, NY 10065, USA e-mail: juluruk@mskcc.org

S. He TeraRecon, Inc., San Mateo, CA, USA

## J. Eng

Johns Hopkins University School of Medicine, Baltimore, MD, USA

# K. Juluru

Memorial Sloan Kettering Cancer Center, 1275 York Ave., Box 29, New York, NY 10065, USA

exists due to geometrical limitations. This can lead to false interpretations of nodule growth or stability that could affect clinical management. The variance is most pronounced at higher slice thicknesses and for small nodule sizes. Measurement error is slice thickness and threshold dependent.

Keywords Chest CT  $\cdot$  Simulation  $\cdot$  Clinical oncology  $\cdot$  Lung neoplasms

# Introduction

Lung nodules detected by computed tomography (CT) often carry an indeterminate diagnosis. As it is impractical to obtain a tissue sample of every nodule, a common strategy to assess a nodule's malignant potential is to follow its size over time. Even when the pathology of a nodule is known, size measurements over time are necessary to assess progression of disease or treatment response. Any nodule can have malignant potential and small changes in diameter can equate to significant changes in volume. In a nodule of 5 mm in diameter, for example, an increase in diameter by only 1.3 mm equates to a doubling of volume. Accurate and precise size assessments, therefore, are essential for the proper diagnosis and determination of treatment response. These measurements are also prerequisites to criterion-based size assessments [1-3] and are the focus of manual, semi-automated, and fully automated techniques that have been proposed to measure both unidimensional size and volume [4, 5].

The problems of nodule size measurement precision and accuracy, often characterized in the literature as "variance" and "bias," respectively, have been well investigated [6]. In a comprehensive review of published studies relevant to lung nodule volumetry by CT, Gavrielides et al. identified multiple technical factors that contribute to size measurement variability [7] (Table 1). However, even when all these factors are

Factor	Description
Scanner type	Number of detectors, size of detectors, tube-gantry geometry
Acquisition settings	Tube current, tube voltage
Reconstruction settings	Slice thickness, slice overlap, kernel, filter
Nodule characteristics	Size, shape, attenuation, uniformity of attenuation, calcification
Nodule surroundings	Presence of adjacent structures such as blood vessels, pleura
Segmentation technique	Manual, semi-automated, automated; variations in algorithms in semi- and fully automated techniques

 Table 1
 Factors contributing to nodule size measurement variability

controlled, variability may still exist due to the problem of partial volume averaging. Depending on the nodule's position relative to the image slice, the number of voxels occupied by the nodule can vary, leading to variance in measured diameter from one exam to the next (Fig. 1).

The purpose of this study is to investigate how much variability exists in nodule diameter measurement due to variations in nodule position relative to image slice. Our investigation seeks to define the lower limit of variability by using a purely geometric model that assumes the ideal condition of a perfectly spherical nodule of uniform density, a background of uniform density, fixed image acquisition parameters, and segmentation based only on voxel density. An understanding of this variability will help to establish future guidelines in determining when nodules are truly stable or have truly changed in size.

## **Materials and Methods**

Throughout this paper, nodule size is assumed to be diameter unless otherwise stated; for example, a "5 mm nodule" is a nodule of 5 mm in diameter.

The software used in this study, Scanner Simulator (version 3.1.1, TeraRecon, Inc., Foster City, CA, USA), is a computer program that runs in the Windows operating system. With a theoretical nodule of known size, position, and density, and with noise-free theoretical slices through that nodule of user-defined thickness and position, partial volume averaging was simulated. If the nodule completely occupied a voxel, the density of the voxel was the density of the nodule. If the voxel was partially occupied by the nodule (partial volume effect), the density of the voxel was proportional to the difference between nodule density and background density, and the fraction of the nodule occupying the voxel. The exact voxel density was determined using a cosine sensitivity profile, as described in prior work [8].

The predominant technique for size measurement is to first determine the number of voxels containing the nodule [7]. The program determined the number of these "positive voxels" based on a user-defined density threshold. The measurement of a maximum diameter, as utilized in response assessment criteria such as RECIST [2], was determined by calculating the maximum axial distance between positive voxels on the slice containing the highest number of positive voxels (Fig. 1). The data were obtained from purely mathematical simulations only and not derived from imaging data.

The software allowed user adjustment of multiple parameters. For this study, scanner field of view (FOV) was set at 30 cm. With a matrix size of  $512 \times 512$ , voxel transverse dimensions were therefore approximately 0.6×0.6 mm. Nodule density was set at 0 HU, and lung density was set at -838 HU, based on results of prior work [9, 10]. Density threshold was set at -419 HU, the halfway point between the lung and nodule densities. Nodule in-plane position was set to "random," meaning that between simulated scans, the axial position of voxels changed randomly. A range of nodule sizes was set from 1 to 10 mm, in 1 mm increments. The experiment was repeated at three different slice thicknesses: 0.75, 3.75, and 5 mm. At each slice thickness, a nodule of a given size was "scanned" 100 times (to simulate 100 consecutively repeated scans on the same patient), each scan with a slice start position that varied by (slice thickness)/100. For a scan with slice thickness of 5 mm, for example, the start position of a slice from one complete "scan" to the next differed by 0.05 mm. Each simulated scan included enough slices to cover the entire nodule. One hundred simulated scans of 10 nodules at three slice thicknesses resulted in 3000 scan simulations.

Average measurement error of each nodule at a given slice thickness was calculated as

$$\frac{\sum_{i=1}^{100} (\text{measured diameter})_i - (\text{true diameter})_i}{100}$$
(1)

where *i* is the scan number. Standard deviations of measured diameters in the 100 scans were calculated for each nodule at each slice thickness. The measured diameter was reported in frequency tables after rounding to the nearest millimeter, a common practice of radiologists. Whereas in the primary analysis, the nodule size varied while the threshold for voxel "positivity" remained constant (at -419 HU), in a sub analysis, nodule size was kept constant (at 5 mm) while varying the threshold for voxel "positivity" from -800 to 0 HU in increments of 100 HU. Analysis of these nine thresholds at three slice thicknesses required an additional 2700 scan simulations.



**Fig. 1** Segmentation of a 10-mm nodule using 5 mm slice thickness, demonstrating variations in partial volume effect due to changes in position of slice to nodule.  $A_{long}$  in longitudinal plane, with slice cutting the upper half of the nodule exactly, only 4 voxels (shaded, with only two shown in this view) are free of partial volume effect.  $A_{trans}$  transverse view of nodule and slice shown in  $A_{long}$  in longitudinal plane, with slice shifted inferiorly relative to nodule, 60 voxels (shaded, with only 10 shown in this view) are free of partial volume effect.  $B_{trans}$  transverse view of nodule and slice shown in  $B_{long}$ , showing 60 voxels (*shaded*) that

are free of partial volume effect. Of voxels that are partially occupied by the nodule, "positivity" is determined by the threshold used. Maximum diameter is determined as maximum distance between "positive" voxels on slice containing highest number of "positive" voxels. When threshold is set very low, voxels are more likely to be positive, leading to over measurement of diameter (*vellow line*). When threshold is set very high, fewer voxels are likely to be positive, leading to under measurement of diameter (*red line*, where threshold=nodule density=0 HU). When threshold is set between these two extremes, values of measured diameter will fall between the measurements of the *yellow line* and *red line* 

# Results

Average error of measured diameter as a function of nodule size at varying slice thicknesses is shown in Fig. 2. At a given slice thickness, average error approaches zero as nodule size increases. For a given nodule size, average error becomes more negative as slice thickness increases. Notably, as nodule diameter increases, there are points at which the absolute value of average error at 3.75 and 5 mm slice thicknesses separately become less than the average error at 0.75 mm slice thickness.

The standard deviation of measured diameter as a function of simulated nodule diameter is shown in Fig. 3, demonstrating that variability increases as slice thickness increases and as nodule size decreases. At very small nodule diameters, variability decreases because the nodule is frequently not detected (diameter=0).



**Fig. 2** Average percent error of measured diameter in 100 scans as a function of nodule size at three slice thicknesses. Each scan varies in start position by (slice thickness)/100 in relation to previous scan. Assumptions: nodule is a perfect sphere, nodule density=0 HU, lung density=-838 HU, voxel positivity threshold=-419 HU, and noise=0



**Fig. 3** Standard deviation of measured diameter in 100 scans as a function of nodule size at three different slice thicknesses. Each scan varies in start position by (slice thickness)/100 in relation to previous scan. Assumptions: nodule is a perfect sphere, nodule density=0 HU, lung density=-838 HU, voxel positivity threshold=-419 HU, noise=0

The frequency of rounded measured nodule diameters for each nodule size is shown in Tables 2, 3, and 4, for slice thicknesses of 0.75, 3.75, and 5 mm, respectively. The tables should be used to appreciate the range and frequency of measurements. For example, Table 4 demonstrates that 100 "scans" of a simulated 5-mm nodule at a slice thickness of 5 mm, with each scan differing in start position of 0.05 mm and keeping all other parameters constant, resulted in rounded diameter measurements of 0 mm—2 times, 1 mm—2 times, 2 mm—7 times, 3 mm—14 times, 4 mm—52 times, and 5 mm—23 times.

Table 2Frequency of rounded nodule diameter measurements in 100scans at slice thickness of 0.75 mm. Each scan varied in start position by(slice thickness)/100 in relation to previous scan. True nodule diameter is

Slice thickness (0.75)

Figure 4 shows the minimum and maximum values of rounded measured diameters of nodules ranging from 1 to 10 mm, in the 100 "scans" differing only in slice start position, at each of the three slice thicknesses. The figure should be used to appreciate the degree of interpretation error that can be introduced, based on the assumptions described in the methodology section (including nodule density, lung density, and voxel threshold) not only when slice thickness is kept constant between multiple scans but also when slice thickness varies between scans on the same patient as what may occur when a CT scan at 5 mm slice thickness performed at one institution is followed with a CT scan at 0.75 mm slice thickness at another institution.

The results of the subanalysis assessing average error of diameter measurement of a simulated 5-mm nodule as a function of threshold are shown in Fig. 5. At the lowest threshold of -800 HU, average error is positive at all slice thicknesses and becomes progressively more negative as threshold approaches 0 HU. At a given threshold, average error increases (becomes more positive) as slice thickness decreases.

### Discussion

The partial volume effect is a well-known and well-described phenomenon in cross-sectional imaging. It is commonly assumed, however, that if it were possible to control all acquisition parameters shown in Table 1 and if all measurements were made using the same measurement tool and/or

shown across the top. Rounded measured diameter on left. Assumptions: nodule is a perfect sphere, nodule density=0 HU, lung density=-838 HU, voxel positivity threshold=-419 HU, and noise=0

Rounded, measured diameter	Nodule diameter										
	1	2	3	4	5	6	7	8	9	10	Grand total
0	93										93
1	7										7
2		100									100
3			100								100
4				45							45
5				55	9						64
6					91	2					93
7						98	1				99
8							99	2			101
9								98	4		102
10									96		96
11										100	100
Grand total	100	100	100	100	100	100	100	100	100	100	1000

**Table 3**Frequency of rounded nodule diameter measurements in 100scans at slice thickness of 3.75 mm. Each scan varied in start position by(slice thickness)/100 in relation to previous scan. True nodule diameter is

shown across the top. Rounded measured diameter on left. Assumptions: nodule is a perfect sphere, nodule density=0 HU, lung density= -838 HU, voxel positivity threshold=-419 HU, and noise=0

### Slice thickness (3.75)

	Nodule diameter										
Rounded, measured diameter	1	2	3	4	5	6	7	8	9	10	Grand total
0	100	100	27								227
1			9	1							10
2			59	11							70
3			5	51							56
4				37	28						65
5					72	17					89
6						83	7				90
7							93				93
8								100			100
9									100		100
10										100	100
Grand total	100	100	100	100	100	100	100	100	100	100	1000

algorithm, that there would be no variability in these measurements. This study demonstrates that there exists an important source of variability in cross-sectional imaging associated with the relative position of nodule to slice.

Figures 2 and 3 support the face validity of the simulation results and also demonstrate some interesting phenomena. As nodule diameter increases, average error of measurement approaches 0 at all slice thicknesses, as expected. However, there is a persistent over measurement at 0.75 mm slice thickness,

whereas at 3.75 mm slice thickness, average error is predominantly negative but crosses 0 and becomes positive above a nodule diameter of about 8 mm. This phenomenon occurs because the algorithm used in the study, similar to commercial algorithms and human observers, measures diameter on the slice containing the greatest number of positive voxels. At small slice thicknesses (such as 0.75 mm), small portions of a nodule's edge that partially fill a voxel are more likely to cause the density of a voxel to achieve the set threshold, thereby

Table 4	Frequency of rounded nodule diameter measurements in	100
scans a	t slice thickness of 5.0 mm. Each scan varied in start position	ı by
(slice th	nickness)/100 in relation to previous scan. True nodule diameter	er is

shown across the top. Rounded measured diameter on left. Assumptions: nodule is a perfect sphere, nodule density=0 HU, lung density=-838 HU, voxel positivity threshold=-419 HU, and noise=0

Slice thickness (5)

Rounded, measured diameter	Nodule diameter										
	1	2	3	4	5	6	7	8	9	10	Grand total
0	100	100	65	25	2						292
1			25	3	2						30
2			10	13	7						30
3				59	14						73
4					52	13					65
5					23	31					54
6						56	35				91
7							65	21			86
8								79	21		100
9									79	11	90
10										89	89
Grand Total	100	100	100	100	100	100	100	100	100	100	1000



Fig. 4 Ranges of diameter measurements, rounded to nearest mm, in 100 scans of nodules ranging in size from 1 to 10 mm at three slice thicknesses. Each scan varies in start position by (slice thickness)/100 in relation to previous scan. Assumptions: nodule is a perfect sphere, nodule density=0 HU, lung density=-838 HU, voxel positivity threshold=-419 HU, and noise=0

leading to over measurement of diameter (Fig. 1). At larger slice thicknesses, the reverse is true, thereby leading to under measurement. The phenomenon is threshold dependent, as confirmed in Fig. 5. The lower (more negative) the threshold, the higher the likelihood that a voxel will be considered positive, causing an over measurement error. As the threshold increases, the number of positive voxels decreases, causing an under measurement error. Importantly, Figs. 3 and 5 show that reduction in slice thickness does not necessarily correlate to a reduction in measurement error; the true size of the object being measured and the threshold used to segment voxels, whether by a computer algorithm or the human eye, play important parts in measurement error independent of slice



**Fig. 5** Average error of measured diameter in 100 scans as a function of voxel positivity threshold of a 5-mm nodule at three different slice thicknesses. Each scan varies in start position by (slice thickness)/100 in relation to previous scan. Assumptions: nodule is a perfect sphere, nodule density=0 HU, lung density=-838 HU, and noise=0

thickness. Using small slice thicknesses, however, keeps variability in measurements to a minimum (Fig. 3).

Management differences can occur between scans performed at a given slice thickness, and between scans performed at different slice thicknesses. On scans performed at 3.75 mm slice thickness, for example, Fig. 4 shows that the rounded, measured diameters of a 5-mm nodule vary from 4 to 5 mm. Based on Fleischner Society guidelines in a low-risk patient, a 4-mm nodule would be managed with no further follow-up, whereas a 5-mm nodule would be managed with a 12-month follow-up CT [1]. If, by chance, the patient with a measured 4-mm nodule happened to have a follow-up CT scan at the same slice thickness, it is possible that at this later timepoint, the nodule is measured to be 5 mm, leading to an interpretation of significant growth despite no actual growth.

Interpretation errors can also occur in the form of describing stability when in fact a nodule has changed in size. Figure 4 shows, for example, that on scans performed at 3.75 mm slice thickness, the maximum rounded measured diameter of a 4mm nodule is 4 mm. At the same slice thickness, the minimum rounded measured diameter of a 5-mm nodule is also 4 mm. Therefore, even when a nodule has truly increased in size from 4 to 5 mm (a doubling of volume), it is possible that the rounded measured diameters could be the same between initial and follow-up scans. A patient with a CT report describing this erroneous stability would likely have no further followup, despite having a growing, potentially malignant nodule.

Changes in slice thickness between baseline and follow-up CTs can also lead to interpretation errors. Figure 4 shows that a 4-mm nodule, when measured on a scan using 5 mm slices, will have a maximum rounded measured diameter of 3 mm. The same 4-mm nodule, when measured on a scan using 0.75 mm slices, will have a minimum measured diameter of 4 mm. A change in slice thickness from 5 to 0.75 mm, therefore, will consistently lead to an interpretation of *at least* a 1-mm growth of an unchanged 4-mm nodule, based on the assumptions made in this study. A patient with a CT report describing this erroneous growth may undergo further mental anguish and unnecessary additional testing.

This study has some important limitations. As a mathematical experiment only, it does not account for a variety of realworld factors. However, the very point of the experiment was to understand the inherent variability in measurements when all the other acquisition and measurement factors are controlled. The study does not account for image noise or irregularities in the shape of a nodule. It does not account for the presence or absence of adjacent structures that could limit the ability to perform the measurements. The distribution of nodule measurements are based on the assumption that the lung is at a uniform density of -838 HU, and the nodule is at a uniform density of 0 HU, yet certainly lung and nodule densities vary from patient to patient. The study uses a fixed threshold of -419 HU to determine whether or not a voxel includes a nodule. The threshold used by commercial algorithms and the additional image processing that is performed by these algorithms is usually proprietary and despite several attempts to contact commercial vendors, this information could not be obtained. Because of this lack of standardization, using different algorithms to follow the same nodule will likely introduce error in measurement.

Nevertheless, the study makes a first attempt to quantify a component of variability that has been under described in the literature—the variability associated with the position of nodule to slice. Most technical assumptions were conservative. The simulation used a very strict model for segmentation. Only available voxels were used for calculations. There was no interpolation or subsampling of data as these techniques do not change the inherent resolution of a system.

So called "coffee-break" experiments have documented the variability in measurement of nodule size on very-short-interval repeat scans [11]. This variability has often been attributed to generalized "system" deficiencies (including deficiencies in the measurement algorithm used). This study shows that even when using an image acquisition and measurement system that is perfect in every other way, there are inherent limitations to accuracy and precision of size measurement in cross-sectional imaging due to geometry and the discrete nature in which nodules are segmented. Importantly, radiologists should exercise great caution when describing either stability or small changes in small lung nodules, notably at higher slice thicknesses. Radiologists should also use caution when describing stability or changes in nodule size when comparing nodules seen on scans of different slice thicknesses.

One potential future study could investigate the variability of measurement algorithms by creating true images with nodules of known size at different nodule-slice positions (as opposed to this study which was purely mathematical) and applying the algorithms for measurement. Although the use of nodule axial diameter is a common practice when attempting to assess for stability or change, future studies should also direct attention, as some have proposed [5, 7, 12, 13], to the inherent variability in volumetric measurement also associated with nodule-slice position.

## Conclusion

Cross-sectional imaging can be a valuable tool in the assessment of nodule size for the purposes of determining malignancy potential or treatment response [14, 15]. There is inherent variability in measurement of nodule diameter that is due to the position of nodule to slice and the resulting changes in partial volume effect. Variability is greatest for small nodules and for higher slice thicknesses used in cross-sectional imaging. This variability could result in clinically significant false interpretations of growth or stability. Radiologists should be aware of these limitations when using cross-sectional imaging for assessment of disease progression or treatment response.

## References

- MacMahon H, Austin JH, Gamsu G, et al: Guidelines for management of small pulmonary nodules detected on CT scans: a statement from the Fleischner Society. Radiology 237(2):395–400, 2005
- Eisenhauer EA, Therasse P, Bogaerts J, et al: New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer 45(2):228–247, 2009
- Organization WH. WHO (World Health Organization, Genf): Handbook for Reporting Results of Cancer Treatment: WHO, 1979
- Marten K, Engelke C: Computer-aided detection and automated CT volumetry of pulmonary nodules. Eur Radiol 17(4):888–901, 2007
- Buckler AJ, Mulshine JL, Gottlieb R, Zhao B, Mozley PD, Schwartz L: The use of volumetric CT as an imaging biomarker in lung cancer. Acad Radiol 17(1):100–106, 2010
- Gavrielides MA, Kinnard LM, Myers KJ, et al: A resource for the assessment of lung nodule size estimation methods: database of thoracic CT scans of an anthropomorphic phantom. Opt Express 18(14):15244–15255, 2010
- Gavrielides MA, Kinnard LM, Myers KJ, Petrick N: Noncalcified lung nodules: volumetric assessment with thoracic CT. Radiology 251(1):26–37, 2009
- McNitt-Gray M: MO-A-ValB-01: Tradeoffs in Image Quality and Radiation Dose for CT. Med Phys 33:2154, 2006
- Xu DM, van Klaveren RJ, de Bock GH, et al: Role of baseline nodule density and changes in density and nodule features in the discrimination between benign and malignant solid indeterminate pulmonary nodules. Eur J Radiol 70(3):492–498, 2009
- Sverzellati N, Randi G, Spagnolo P, et al: Increased mean lung density: another independent predictor of lung cancer? Eur J Radiol 82(8):1325–1331, 2013
- Mozley PD, Bendtsen C, Zhao B, et al: Measurement of tumor volumes improves RECIST-based response assessments in advanced lung cancer. Transl Oncol 5(1):19–25, 2012
- Goodman LR, Gulsun M, Washington L, Nagy PG, Piacsek KL: Inherent variability of CT lung nodule measurements in vivo using semiautomated volumetric measurements. AJR Am J Roentgenol 186(4):989–994, 2006
- Wormanns D, Kohl G, Klotz E, et al: Volumetric measurements of pulmonary nodules at multi-row detector CT: in vivo reproducibility. Eur Radiol 14(1):86–92, 2004
- Xu DM, Gietema H, de Koning H, et al: Nodule management protocol of the NELSON randomised lung cancer screening trial. Lung Cancer 54(2):177–184, 2006
- Xu DM, van der Zaag-Loonen HJ, Oudkerk M, et al: Smooth or attached solid indeterminate nodules detected at baseline CT screening in the NELSON study: cancer risk during 1 year of follow-up. Radiology 250(1):264–272, 2009