Operator-dependent reproducibility of size measurements of small phantoms and lung nodules examined with low dose thin-section

Operator-Dependent Reproducibility of Size Measurements of Small Phantoms and Lung Nodules Examined With Low-Dose Thin-Section Computed Tomography

Giulia Picozzi, MD, Stefano Diciotti, PhD,† Massimo Falchini, MD,* Silvia Foresti, MD,* Francesca Gallesi, MD,* Edoardo Cavigli, MD,* Lorenzo Livi, MD,‡ Natale Villari, MD,* and Mario Mascalchi, MD, PhD**

Objective: We sought to assess the reproducibility of size measurements of small lung nodules examined with low-dose thin-section computed tomography (LDTSCT).

Materials and Methods: Three radiologists measured volume with a semiautomatic tool and diameters manually of 20 (equivalent diameter range, 5.3–11 mm) phantom nodules and 37 (mean diameter range, 5–8.5 mm) lung nodules in subjects undergoing LDTSCT.

Results: In phantoms, the worst 95% limits of agreement (95% LA) for volume were −3.0% and 3.0% within operator and −3.1% and 2.8% between operators. The coefficient of repeatability (CR) for diameter ranged between 0.51 and 0.67 mm within operator and the 95% LA were from −0.71 to 0.71 mm between operators. In nodules, the worst intraoperator 95% LA for volume were −14.4% and 17.6% within operator and −13.1% and 14.2% between operators. The CR for diameter ranged between 0.48 and 0.73 mm within operator and the 95% LA were from −1.16 to 1.16 mm between operators.

Conclusion: Operator-dependent variability of size measurements of small nodules examined with LDTSCT is not negligible and should be considered in lung cancer-screening studies.

Key Words: phantoms, lung nodules, low-dose CT, reproducibility, computer diagnostic aids

(*Invest Radiol* 2006;41: 831–839)

Computation of the volume of lung nodules with dedicated software decreases the variability of 2-dimensional (2D) measurements and has the potential to increase the sensitivity of nodule growth assessment with computed tomography (CT).[1–7] This is relevant in general for nodules detected in nonscreening CT[8] but in particular in the context of lung cancer screening, in which small indeterminate nodules at baseline or annual repeat are frequently re-examined after a few months follow-up and evidence of nodule growth represents the fundamental criterion for clinical management.[9] Although initial protocols for lung cancer screening with CT used whole-lung low-dose, namely low tube-current, thick sections that were supplemented with standard dose thin-section targeted acquisitions to assess nodule size,[10] current screening protocols recommend whole lung low-dose thin-section CT (LDTSCT).[11] The possibility to measure nodule size directly on LDTSCT is attractive but raises concerns about quality and robustness of measurements in images with a decreased signal to noise ratio as the result of low tube-current. Moreover, previous studies testing software for volumetric measurements of lung nodules were predominantly performed in CT scans acquired with standard tube-current.[1,2,5] As a preliminary step before adoption of a custom-made software for lung nodule volume computation in a clinical trial for lung cancer screening with LDTSCT, we evaluated the reproducibility of volume and 2D size measurements of small nodule phantoms and indeterminate lung nodules identified in subjects undergoing LDTSCT.

MATERIALS AND METHODS

Nodule Phantoms

The set of nodule phantoms was constituted by a sample of 3 types of phantoms similar in density to lung nodules (Fig. 1) on which the software was validated. Type A were 10 silicone synthetic nodules (mean density approximately 115 Hounsfield Units [HU]) with irregular shape and known volume (mean volume, 268.8 mm^3; range, 75.9–688.6 mm^3; mean equivalent diameter, 8.0 mm; range, 5.3–11.0 mm). Type B were 5 synthetic silicone phantoms simulating nodules identical in size to type A phantoms adjacent to different pulmonary structures such as vessels. Type C (mean volume, 300.6 mm^3; range, 196.4–445.9 mm^3; mean equivalent diameter, 8.3 mm; range, 7.2–9.5 mm) were 5 small deformable phantoms made of silicone rubber mixed with policarbonate microspheres (mean density approximately 50 HU). All nodules were embedded in a marjoram

Received June 3, 2006, and accepted for publication, after revision, July 31, 2006.

From the *Radiodiagnostic and ‡Radiation Therapy Sections, Department of Clinical Physiopathology, and †Department of Electronics and Telecommunications, University of Florence, Florence, Italy.

Supported by the Health Department of the Tuscany Region, Italy and by the Ministry of Instruction, University and Scientific Research of Italy, Grant No. 20030680017.

Reprints: Mario Mascalchi, MD, PhD, Radiodiagnostic Section, Department of Clinical Physiopathology, University of Florence, Viale Morgagni 85, Florence Italy. E-mail: m.mascalchi@dfc.unifi.it.

ISSN: 0020-9996/06/4111-0831

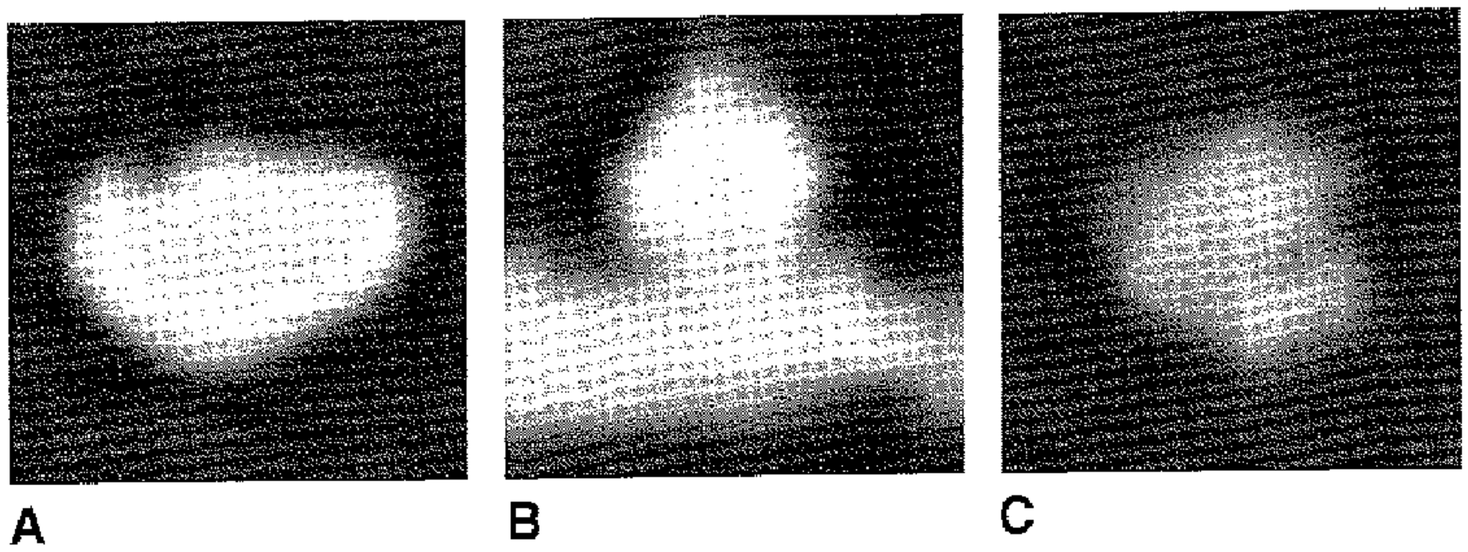


FIGURE 1. CT appearance of type A (A), type B (B) and type C (C) phantoms. The images are displayed with a standard lung window settings (window level = −500 HU, window width = 1500 HU). Type A phantoms are well-circumscribed silicone synthetic nodules with irregular shape. Type B are silicone synthetic phantoms simulating nodules adjacent to different pulmonary structures. Type C are deformable nodule phantoms. All nodules are embedded in a marjoram background.

(mean density, −865 HU) background to reproduce the texture of the lung parenchyma in the CT scans.

Lung Nodules

The set of lung nodules studied was constituted by 37 indeterminate solid nodules exceeding 5 mm in mean diameter (mean, 6.3 ± 1.1 mm; range, 5.0–8.5 mm as determined by caliper measurements on the main CT console by the same operator; see CT scan acquisition protocol in this article) and correctly segmented which were observed in the first consecutive 35 subjects (2 subjects had 2 nodules each) showing such nodules in the active arm of the Italung-CT trial at baseline examination.[12] Twenty-one subjects were examined on a single-slice CT (SSCT) scanner, whereas 14 were scanned on a multislice CT (MSCT) scanner (see CT Scan Acquisition Protocols). Fourteen nodules were well circumscribed, 15 were juxtavascular, and 8 had a pleural tail.[13] Since 4 additional juxtavascular nodules were judged not to be correctly segmented by an experienced supervisor, the segmentation failure rate was 4 of 41 (9.8%). The nodules were all solid, noncalcified, and their shape was smooth in 16, lobular in 2, irregular in 14, and spiculated in 5.

CT Scan Acquisition Protocols

Lung nodules were examined with LDTSCT on a SSCT Somatom Plus scanner (Siemens Erlangen, Germany) with 1 second rotation time (23 lung nodules) or a MSCT Somatom Plus 4 Volume Zoom scanner (Siemens Erlangen, Germany) with 0.5 seconds rotation time and 4 rows of detectors (14 lung nodules). The low dose thin-section technique included 3 mm collimation, pitch 2, reconstruction 1.5 mm, 140 kV, 43 mAs, AB91 reconstruction filter, 2 breath-holds with SSCT scanner and 1.25 mm collimation, pitch 7, reconstruction 1 mm, 140 kV and 40 mAs, B70f reconstruction filter, 1 breath-hold with MSCT scanner.

All the 20 nodule phantoms were examined both on the same SSCT and the MSCT scanners used for lung cancer screening with the same scanning parameters (see the section "Lung Nodules"). For nodule phantoms, a fixed field of view of 340 mm was chosen which is similar to that used in subjects undergoing screening with LDTSCT for lung cancer screening.

Software

A semiautomatic software for computation of the volume of lung nodules in CT scans was developed by one of the authors (S.D.) at the Department of Electronics and Telecommunications of the University of Florence. The software was validated in phantoms using standard and low tube-current thin-section acquisition techniques (Root Mean Square Error in phantoms of type A and B was consistently less than 6.6%). The operator selects the nodule of interest in the spiral CT scans by using a custom computer user interface (step 1). Because we were interested in assessing the volume of small nodules (diameter between 5 and 10 mm) a cubic volume of interest (VOI) of 25 × 25 × 25 mm^3 is extracted for successive processing (step 2). The VOI is supersampled with tri-linear interpolation to obtain an isotropic voxel and to reduce partial volume effects (step 3).[13] The first processing step is a focus of attention stage for the identification of pulmonary structures, both nodular and non-nodular, such as vessels and pulmonary fissures (step 4). The results of the focus of attention stage are shown to the operator for visual inspection to classify each identified structure as nodular, non-nodular or noise (step 5). The location of the pulmonary structures was kept into account by the following stage, the 3D region-growing segmentation algorithm, to improve the segmentation of juxtavascular nodules (step 6). The algorithm is based on both a thresh-holding technique and the geodesic influence zones concept and the volume of the nodule is thus computed by using the voxel counting method (step 7). Because no pleura identification was employed in the focus of attention stage, juxtapleural nodules were excluded by our analysis.

Reproducibility

Three radiologists with at least 4 years of thoracic imaging experience (G.P., S.F., and F.G.) evaluated the 2 sets of nodules working independently. Each radiologist measured twice the volume and 3 times the diameters of 20 nodule phantoms scanned twice (in 2 scanners) and 37 lung nodules. The nodule measurement sessions were separated by at least 1 week. A total of 240 volume nodule phantom measurements (80 for operator) and of 222 volume lung nodules measurements (74 for operator) were obtained.

According to the I- ELCAP protocol,[11] nodule diameter is the average of length and width. Length is measured in a single CT image, which shows the maximum length. The width is the largest perpendicular to the length and is measured in the same CT image. A total of 360 2D nodule phantom measurements (120 for operator) and of 333 2D lung nodules measurements (111 for operator) were obtained.

TABLE 1. Volume, Mean Diameter, and Maximum Diameter Measurements of Type A Phantoms Obtained From Scans Acquired on SSCT and MSCT

	SSCT	MSCT	*P* Wilcoxon Matched Pair Test
Volume (mm^3)			0.88
Mean ± SD	431.9 ± 119.3	432.9 ± 117.3	
Range	266.6–696.5	268.0–691.5	
Mean diameter (mm)			0.50
Mean ± SD	9.0 ± 1.1	9.1 ± 1.0	
Range	6.8–11.7	7.1–11.6	
Maximum diameter (mm)			<0.01
Mean ± SD	9.7 ± 1.3	10.2 ± 1.1	
Range	7.8–12.2	7.8–12.1	

Only the locations of lung nodules to be measured with reference to the segmental anatomy of the lung was provided to the operators. In particular, they did not receive or record information about the slice on which to perform the 2D measurement or start the 3D assessment.

Statistical Analysis

To assess if the size measurements were influenced by scan acquisition on the SSCT or MSCT scanner, the volume and diameters of type A phantoms were compared using the 2-tailed Wilcoxon matched paired test with a significance level of $P < 0.05$. In the case of a significant difference a one-tailed test was used to investigate whether one data set was greater than the other. Volumes and mean and maximum diameters measurements of type A phantoms of each radiologist taken on CT scans from SSCT scanner were compared with the same measurements of the same radiologist taken on CT scans from MSCT scanner.

To evaluate operator-dependent reproducibility, we preliminarily evaluated for each operator's data set the correlation between standard deviation (SD) of the measurement and phantom or nodule size using the Kendall τ test with a $P < 0.05$. If such a correlation was not observed, we used the repeatability coefficient (CR), as defined by the British Standards Institution,[14] to measure the intraoperator reproducibility and the Rousson method[15] [95% limits of agreement (95% LA) of the difference between measurements performed by operators] to measure the interoperator agreement. In case of a significant relationship between SD and magnitude of the measurement, log-transformed data were employed and the 95% LA according to the Bland and Altman method was adopted to measure the intra and interoperator reproducibility.[16–18] Finally, to graphically assess the degree of agreement, the mean measurements of all possible operator's pairs and the corresponding lines of equality plot were drawn.

RESULTS

SSCT Versus MSCT Size Measurements

Table 1 shows the volume and mean diameter and maximum diameter measurements of type A phantoms obtained from scans acquired on SSCT and MSCT. The Wilcoxon matched pair test showed that differences in measurements performed on SSCT and MSCT scans were not significant for volume ($P = 0.88$) and mean diameter ($P = 0.50$), whereas they were significant ($P < 0.01$) for maximum diameters. The one-tailed test revealed that maximum diameters were significantly ($P < 0.01$) greater for measurements taken on MSCT scanner as compared with SSCT scanner. In particular, the mean of the differences were 0.55 mm and the median of the differences 0.20 mm.

Operator-Dependent Reproducibility

Table 2 details the results of the Kendall τ test for volume and 2D measurements in nodule phantoms and lung nodules. A significant correlation between SD and magnitude of the nodule phantom volume measurements was observed for one operator, but neither for the other operators nor for the 3 operators combined. A significant correlation between SD and magnitude of lung nodule volume measurements was observed for 1 operator and for the 3 operators combined. No correlation between SD and magnitude of the 2D measurements was observed for any operator and for all 3 operators combined in the case of nodule phantoms and lung nodules.

The intraoperator and interoperator 95% LA for volume measurements in phantoms and lung nodules are detailed in Table 3. In the case of nodule phantoms the intraoperator 95% LA for volume ranged from −0.3% and 0.3% to −3.0% and 3.0% and the interoperator 95% LA ranged from −2.0% and 1.9% to −3.1% and 2.8% (Fig. 2). In the case of lung

TABLE 2. The Kendall τ Correlation Coefficients Between SD and Magnitude of the Measurement in the Case of 2D (Maximum and Mean Diameter) and 3D (Volume) Measurements in Nodule Phantoms and Lung Nodules Measured by 3 Operators

	Nodule Phantoms			Lung Nodules		
Operator	Maximum Diameter	Mean Diameter	Volume	Maximum Diameter	Mean Diameter	Volume
1	0.001	−0.05	0.26*	−0.09	−0.05	0.01
2	0.05	0.05	−0.15	0.06	0.06	0.29*
3	0.12	0.22	−0.21	−0.03	0.03	0.17
All	0.17	0.08	−0.04	0.15	0.14	0.26*

*$P < 0.05$.

TABLE 3. Intraoperator and Interoperator Mean Difference and 95% Limits of Agreement for Volume Measurements

	Nodule Phantoms			Lung Nodules		
Operator	Mean Difference	95% Limits of Agreement		Mean Difference	95% Limits of Agreement	
Intraoperator						
1	0	−0.3%	0.3%	1.6%	−14.4%	17.6%
2	0	−3.0%	3.0%	0.6%	−11.7%	14.6%
3	−0.2%	−1.9%	1.6%	0	−12.4%	12.4%
Interoperator						
1–2	−0.2%	−3.1%	2.8%	−0.2%	−11.2%	12.2%
1–3	−0.1%	−2.0%	1.9%	−0.3%	−13.1%	14.2%
2–3	0.2%	−2.4%	2.7%	−0.2%	−11.4%	12.4%

nodules the intraoperator 95% LA for volume ranged from −12.4% and 12.4% to −14.4% and 17.6% and the interoperator 95% LA ranged from −11.2% and 12.2% to −13.1% and 14.2% (Fig. 3).

The intraoperator CR and interoperator 95% limits of agreement for nodule phantoms and lung nodules are detailed in Table 4 for mean diameter measurements and in Table 5 for maximum diameter measurements. The intraoperator CR for mean diameter in phantoms ranged between 0.51 mm and 0.67 mm and the interoperator 95% LA ranged from −0.71 mm to 0.71 mm (Fig. 4). The intraoperator CR for maximum diameter in phantoms ranged between 0.67 mm and 0.91 mm and the interoperator 95% LA ranged from −0.91 to 0.91 mm. The intraoperator CR for mean diameter in lung nodules ranged between 0.48 mm and 0.73 mm and the interoperator 95% LA ranged from −1.16 mm to 1.16 mm (Fig. 5). The intraoperator CR for maximum diameter in lung nodules ranged between 0.66 mm and 1.22 mm and the interoperator 95% LA ranged from −1.46 mm to 1.46 mm. Overall, greater interoperator reproducibility was observed for volume as compared with mean diameter measurements in the case of nodule phantoms and lung nodules.

DISCUSSION

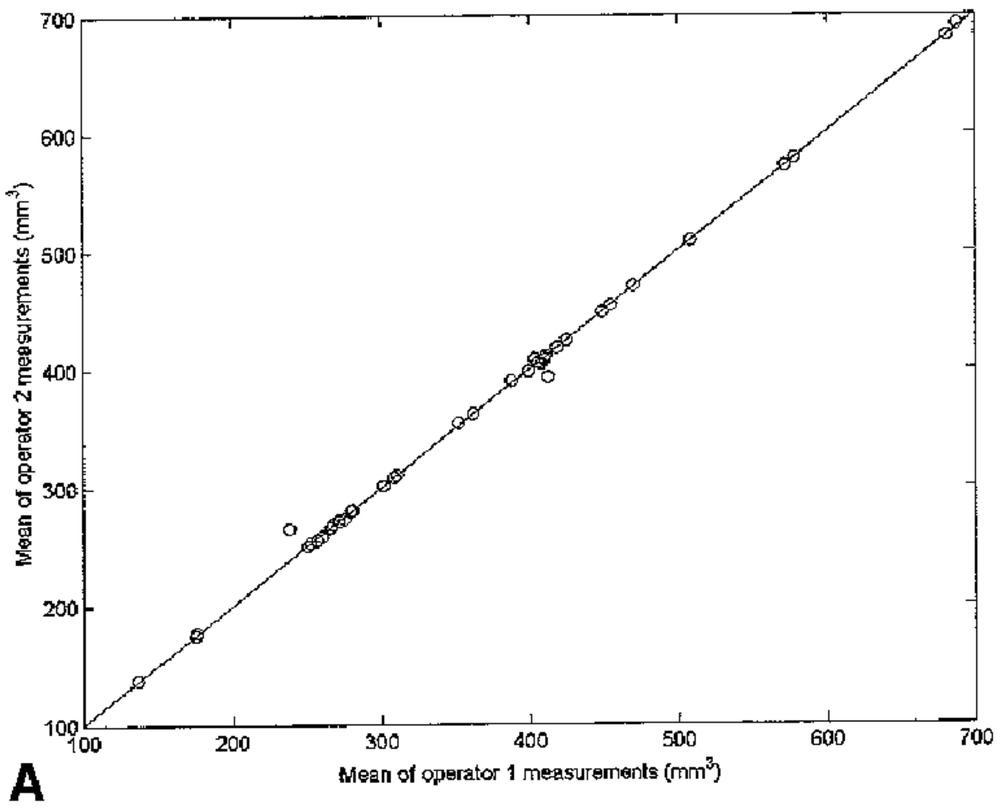
Nodule size assessement in clinical and screening practice is typically performed in 2 dimensions[11,19–21] using the maximum or mean diameter. In a study performed on pulmonary nodules (3–18 mm in diameter) identified in a clinical setting and scanned with standard dose thin-section spiral CT, Revel et al,[22] based on the reproducibility of maximum diameter taken by 3 operators, concluded that the maximum nodule diameter would have to increase by at least 1.7 mm to be confident with a true increase in size of the nodule. In our study, considering the interoperator variability of maximum diameter on lung nodules (95% LA = −1.46 mm to 1.46 mm), we obtained similar results in a screening setting using LDTSCT, which implies a lower signal-to-noise ratio and a worse condition to identify nodule borders. Notably, in both studies, radiologists were not aware of the slice on which to perform the measurement. It is noteworthy that in our study in phantoms assessing interscanner variability only the maximum diameter was significantly different for scans obtained on SSCT or MSCT scanner, presumably reflecting the thinner collimation used on MSCT scanner which implies less partial volume effects. Following the I-ELCAP recommendations,[11] we also measured the mean diameter of the nodules, which was more reproducible (95% LA = −1.16 mm to 1.16 mm) than the maximum diameter.

Low reproducibility of linear measurements,[22,23] the nonspherical shape of most nodules, and the development of multidetector spiral scanning have fuelled interest on volumetric computerized methods to assess lung nodule size. Numerous studies have reported that 3D methods applied to thin-section CT at standard dose enable: (1) the accurate assessment of pulmonary nodule volume[1,3,24] and (2) the reduction of intra and interoperator variability in establishing nodule size.[2,5] In particular, Revel et al[2] evaluated 54 nodules of 5–18 mm in diameter in a clinical setting using thin-section CT at standard dose and reported intraoperator 95% LA of 2.4–3.1% for volume measurements using a fully automatic software. In the case of interoperator disagreement, the widest 95% LA was 8.9% and 7.6% for 2 operators. Wormanns et al[4] in a study with LDTSCT of 151 pulmonary nodules (diameter range 2.2–20.5 mm, mean 7.4 mm) in a clinical setting using an automatic software found an interoperator 95% LA for volume measurements of 5.5 to 6.6% and intraoperator 95% LA of −3.9 to 5.7%.

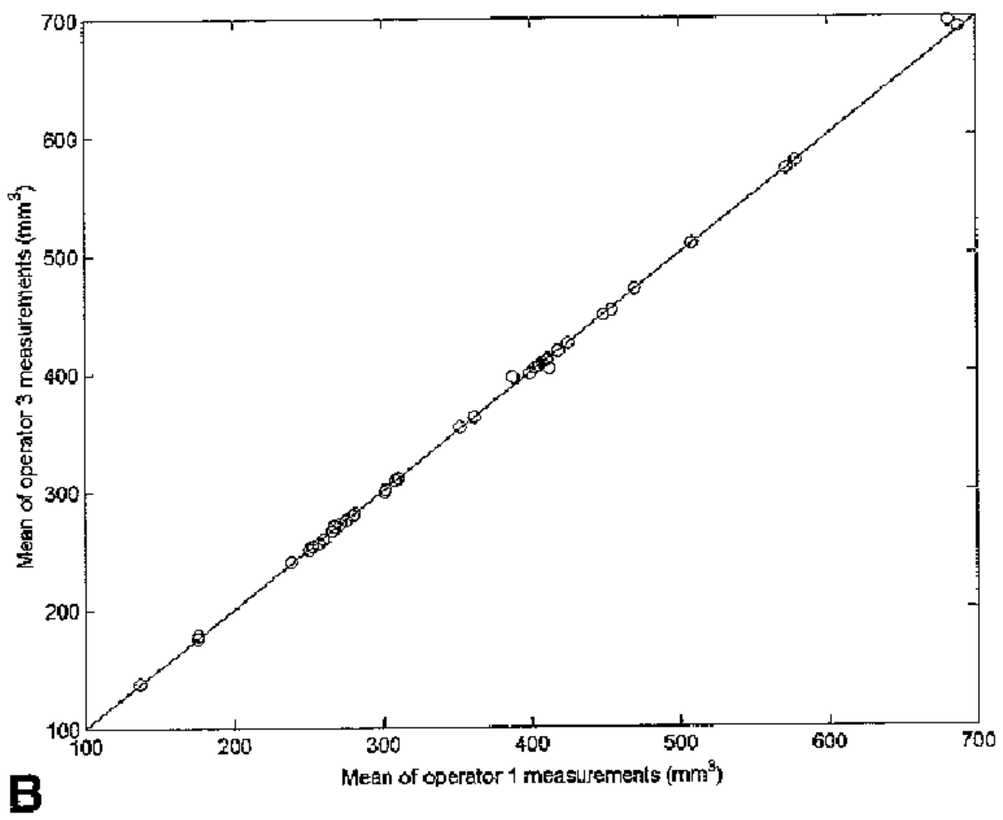
A general problem of the volume measurement approach is the incapability of any software to segment all nodules and notably those juxtapleural with segmentation failure rate ranging between 4% and 12%.[2,5,24] Our semi-automatic software had a segmentation failure rate of 9.8%. This common failure of 3D software makes use of 2D measurements unavoidable for some nodules.

So far, the variability of size measurements in the screening setting has been addressed only in one study,[25] which used targeted (small) field of view and standard tube-current. In our study, we used LDTSCT with a large field-of-view and a semi-automatic custom-made software for volume measurements of small (5.0–8.5 mm in diameter) lung nodules observed in a screening setting.

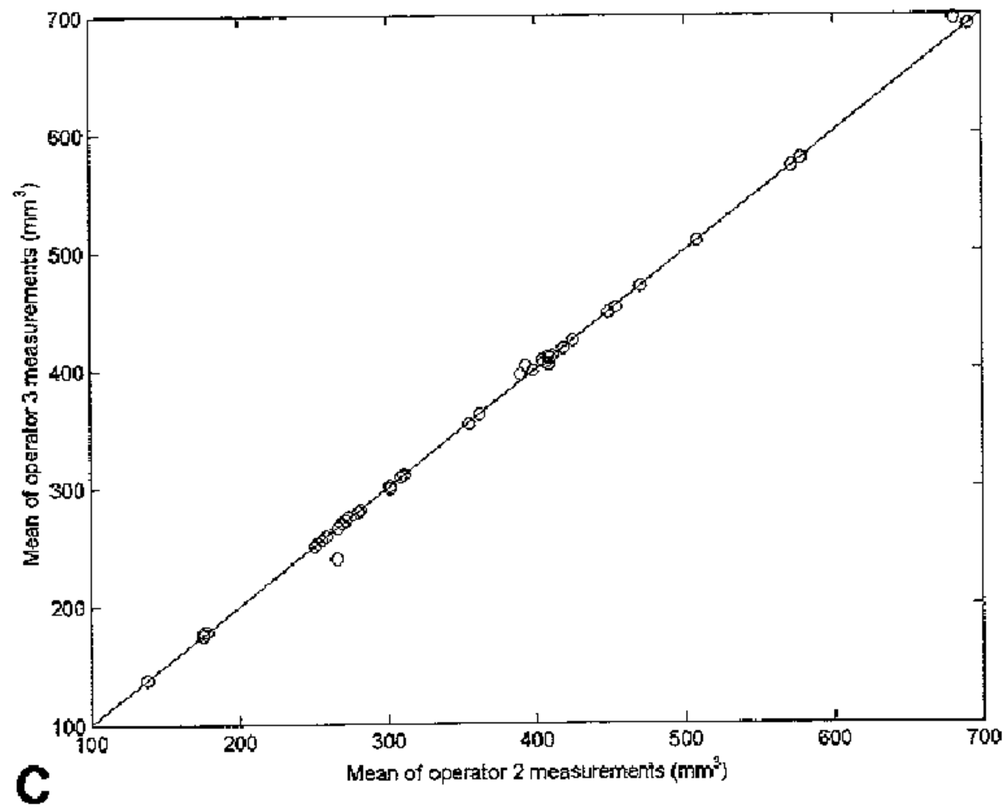
Preliminarily, we assessed the reproducibility of size measurements on small (5.3–11 mm in diameter) nodule phantoms. This represents a simplified setting than that of



A

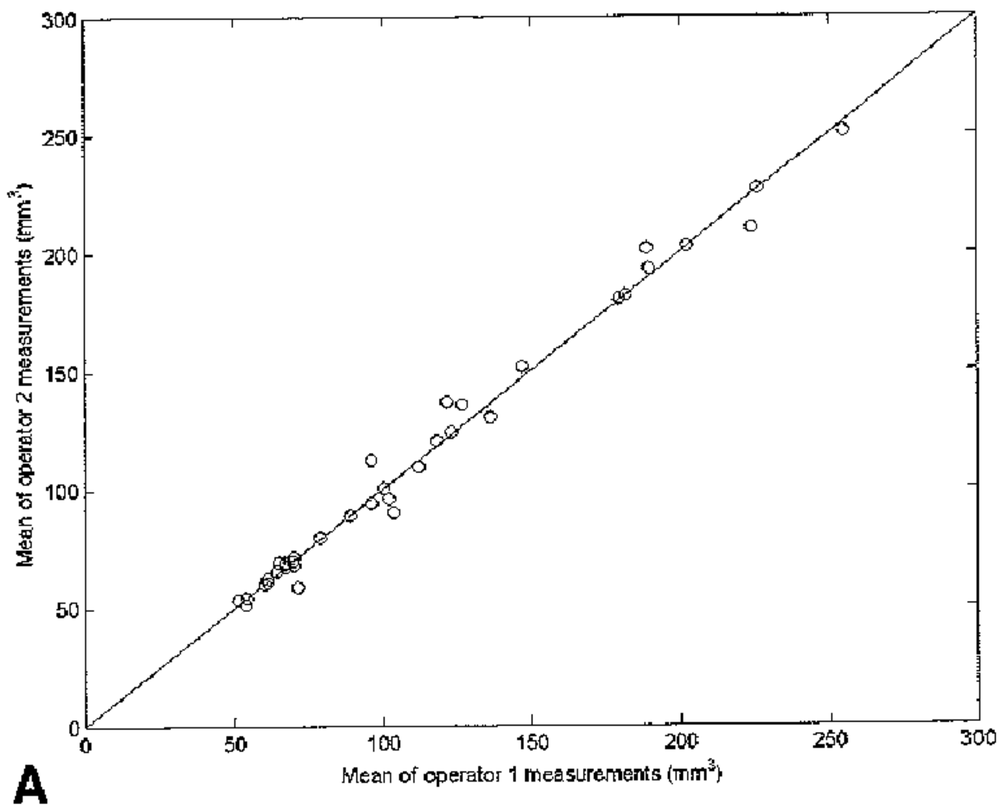


B

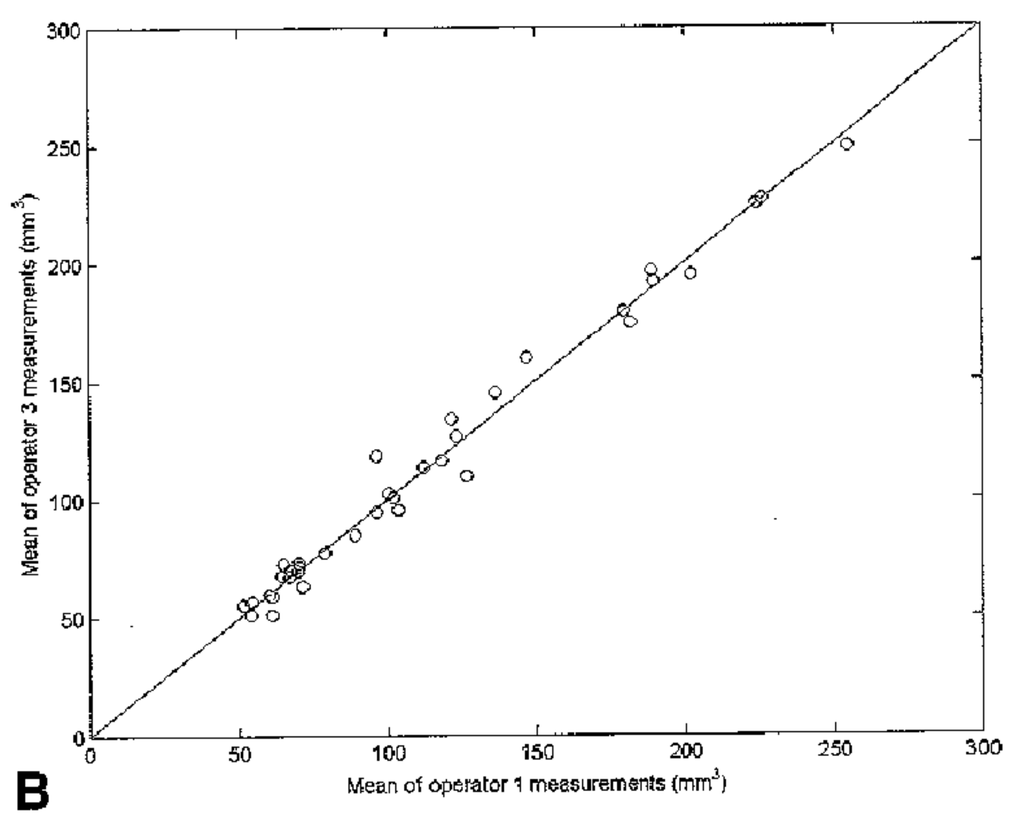


C

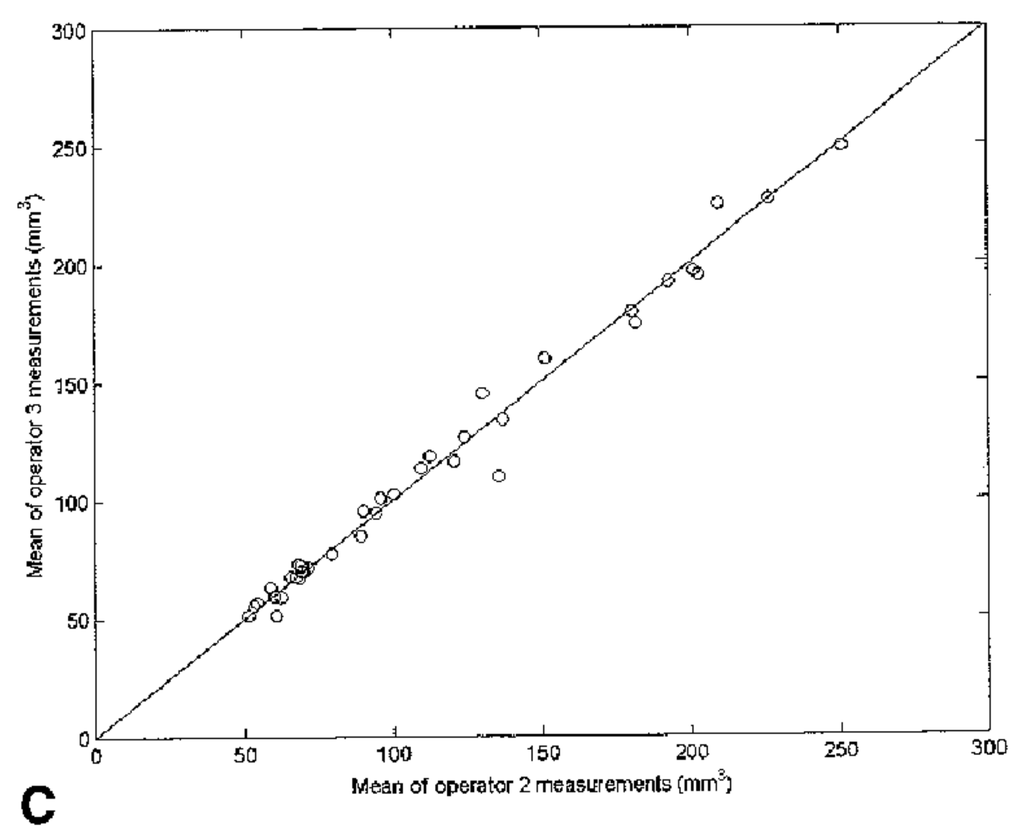
FIGURE 2. A–C, Scatter plots of the mean volume measurements of nodule phantoms: measurements of operator 1 versus 2 (A); measurements of operator 1 versus 3 (B); measurements of operator 2 versus 3 (C). There was excellent agreement, in fact all points lie along the line of equality.



A



B



C

FIGURE 3. Scatter plots of the mean volume measurements of lung nodules: measurements of operator 1 versus 2 (A); measurements of operator 1 versus 3 (B); measurements of operator 2 versus 3 (C). There was good agreement.

TABLE 4. Intraoperator Coefficient of Repeatibility (2.77 × SD With a 5% Error Rate) and Interoperator 95% Limits of Agreement for Mean Diameter Measurements

	Nodule Phantoms			Lung Nodules		
Intraoperator						
Operator	Coefficient of repeatability (mm)			Coefficient of repeatability (mm)		
1	0.67			0.48		
2	0.56			0.73		
3	0.51			0.63		
Interoperator						
Operator	Mean difference (mm)	95% limits of agreement (mm)		Mean difference (mm)	95% limits of agreement (mm)	
All	0	−0.71	0.71	0	−1.16	1.16

TABLE 5. Intraoperator Coefficient of Repeatibility (2.77 × SD With a 5% Error Rate) and Interoperator 95% Limits of Agreement for Maximum Diameter Measurements

	Nodule Phantoms			Lung Nodules		
Intraoperator						
Operator	Coefficient of repeatability (mm)			Coefficient of repeatability (mm)		
1	0.91			0.66		
2	0.67			1.22		
3	0.71			0.91		
Interoperator						
Operator	Mean difference (mm)	95% limits of agreement (mm)		Mean difference (mm)	95% limits of agreement (mm)	
	0	−0.91	0.91	0	−1.46	1.46

lung nodules examination, the main differences being related to lack of any motion in nodule phantom acquisitions as compared with the vulnerability of lung nodule acquisitions to different types of motion such as imperfect breath-holding and cardiac motion. Not surprisingly, in our study the intra and interoperator 95% LA intervals for volume measurements were far smaller for phantoms as compared with lung nodules which were scanned using identical protocols, in particular collimation, reconstruction filter and tube current. The lower reproducibility of lung nodules size measurements is most likely due to difficulties in segmenting lung nodules which are ill-defined, adjacent to blood vessels as compared with the regular shape of phantoms. Additional factors increasing the variability of lung nodules measurements include cardiac and respiratory motion during acquisition, changes in the position of the nodule relative to adjacent structures and changes in nodule shape.[3,4,25–27]

We measured a 95% LA for volume measurements of lung nodules of −14.4% and 17.6% or less within operator and of −13.1% and 14.2% or less between operators. Possible causes of the lower reproducibility we measured as compared with previous studies in a clinical setting[2,4] include the smaller size of the nodules investigated, the low tube-current and the large field of view techniques and the use of a semiautomatic software.

Segmentation procedures generally are aimed to high accuracy and reproducibility. User interaction enables to introduce the a priori knowledge of the human expert which may improve the algorithm accuracy. However, a source of variability due to the user-dependent phase is added and both the within-operator and the between-operator reproducibility can be lower than in fully automatic procedures.

The reproducibility of size measurements we obtained in a screening setting are of potential interest for lung cancer screening practice in several aspects. In fact indeterminate nodule size and growth are fundamental for clinical decision making in lung cancer screening with CT,[10,11,28] in particular in respect to the cut-off size values arbitrarily recommended to select "significant" indeterminate nodules requiring follow-up and to ascertain true nodule growth at follow-up.

Because the detection of nodules less than 5 mm on screen does not justify immediate work-up but only annual repeat screening[29] many CT screening trials use this cut-off value for nodules found at the baseline examination to be scheduled for short term follow up.[11,13,28] At annual repeat, a cut-off of 3 mm in diameter for new nodules is recommended. Recently, a volume cut-off (50 mL) was introduced in the Nelson study to qualify for "significant" indeterminate nodules at baseline screening.[30] From our data on nodules greater or equal to 5 mm, if an operator measures a mean diameter of 5 mm in a given nodule, she or he has a 95% chance to measure a mean diameter ranging from 4.27 to 5.73 mm when a second calculation is made. However, when the mean diameter measurement of the same nodule is taken by another operator there is a 95% chance that this is in the range between 3.84 and 6.16 mm. This imperfect reproducibility should be taken into account when discrepancies emerge in some of the results, for instance recall rate, of screening programs using similar dimensional cut-off values, especially in multicenter studies.

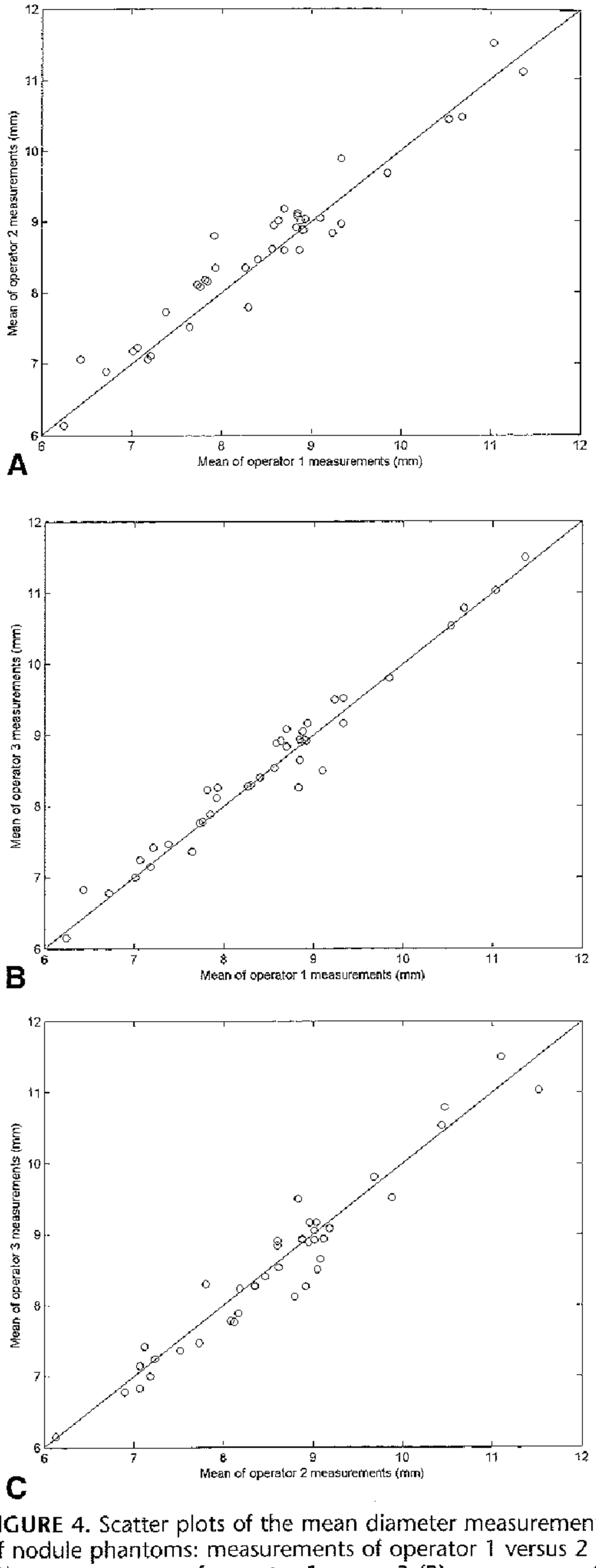


FIGURE 4. Scatter plots of the mean diameter measurements of nodule phantoms: measurements of operator 1 versus 2 (A); measurements of operator 1 versus 3 (B); measurements of operator 2 versus 3 (C). There was fair agreement.

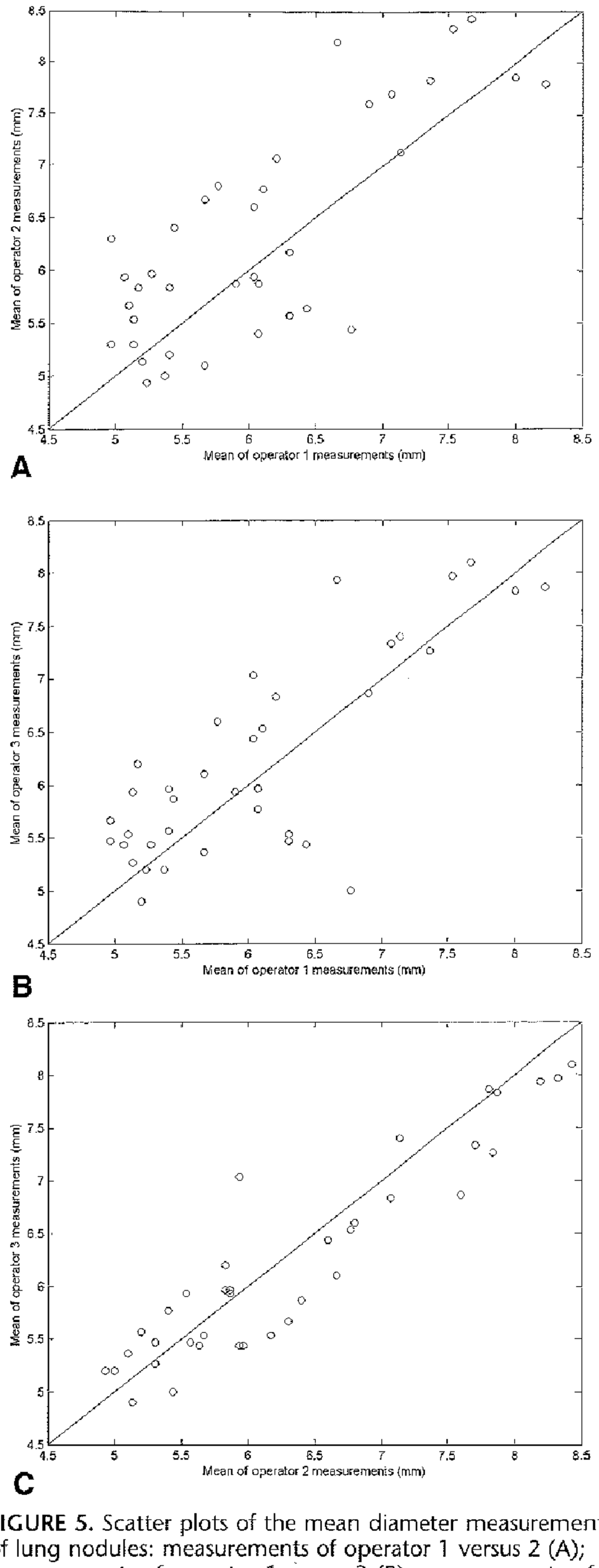


FIGURE 5. Scatter plots of the mean diameter measurements of lung nodules: measurements of operator 1 versus 2 (A); measurements of operator 1 versus 3 (B); measurements of operator 2 versus 3 (C). There was poor agreement, in fact most points lie far from the line of equality.

Nodule growth is one of the most compelling indicators of malignancy, and the primary concern is whether a nodule that appears to have grown in sequential CT examinations has actually grown or the difference in measurements is due to a measurement error. Criteria to establish nodule growth in lung cancer screening with CT have not been established. In line with the RECIST criteria, for which an increase of at least 20% in the maximum diameter (corresponding to an increase of volume of 73%) qualifies for lesion growth/progression of disease[20,21] some authors proposed (EU/US draft guidelines for CT screening for lung cancer, May 2002) an increase of at least 20%, namely 1 mm in mean diameter for a 5-mm nodule, when 2D measurements are used. An increase of at least 25% was proposed to establish nodule growth in case of 3D measurements.[4,30] It is noteworthy that making the cut-off value smaller may be associated with a shorter time to follow-up to establish growth or lack of thereof with the benefit to reduce the subject's anxiety associated to the possibility to harbor a malignant lesion.

In our study the worst case for intraoperator 95% LA for lung nodules volume measurements was 17.6%. This means that any growth of 17.6% or less could not be distinguished, with a 5% error rate, from the measurement error. Moreover, since a volume increase of 100% (lesion doubling) corresponds to an increase of the diameter of only the 26%, volume measurements appear to be more sensitive to reveal a nodule growth respect to a diameter measurement. In fact, an increase of 17.6% in volume corresponds to an increase of only 5.5% in diameter. From our data such diameter growth could not be distinguished from measurement error because CR and the interobserver 95% LA in mean diameter in the worst case were, respectively, 0.73 mm (14.6% for a 5-mm nodule or 7.3% for a 10-mm nodule) and 1.16 mm (23.2% for a 5-mm diameter nodule and 11.6% for a 10-mm diameter nodule). Our data indicate that a 25% in volume increase proposed as a cut-off to establish nodule growth on LDTSCT in lung cancer screening[30] falls beyond the 95% LA of intra and interoperator reproducibility. On the contrary they indicate that 1 mm increase of the mean diameter is within the range of the operator-dependent variability, if measurements from different operators are considered.

We recognize several limitations of our study. First, we considered only indeterminate solid nodules because nonsolid and part-solid nodules pose special challenges to diameter measurements and to the analysis of growth.[31] In addition, juxtapleural nodules were discarded because of the difficulty in nodule segmentation.

Second, the size range of the solid nodules we examined covers only partially that encountered in screening practice. Notably, no nodule less than 5 mm in diameter was considered because our software was validated only for nodule of 5–10 mm diameter size. However, because the variability of the volume measurements using a 3D software is inversely proportional to the size of the nodule,[25] it is expected that larger intervals of the 95% LA than those we measured would apply to nodules smaller than 5 mm.

Third, the data we obtained in phantoms type A indicated a low interscanner variability which enabled us to consider negligible differences in size measurements due to the scanner and perform a cumulative analysis of the lung nodules scanned on SSCT and MSCT. We recognize that this negligibility does not necessarily apply to interscanner variability of other types of phantoms and especially lung nodules.

Fourth, our study we did not consider variations occurring between 2 different CT examinations in the same scanner, namely the scan-rescan variability.[4,32] This is affected by variations due to patient's position, different lung volume due to variable degrees of inspiration, phase of cardiac cycle or presence of microatelectasis,[2,4,5] as well as technical changes such as slice registration and selection which may lead to varying amount of volume averaging. This scan-rescan variability was as high as 20–27% in clinical settings using automatic software and low tube-current techniques[4,24] and represents an additional source of variability in comparing the size of a nodule at initial observation and follow-up. Assessment of scan-rescan variability in the setting of lung cancer screening with LDTSCT in the Italung-CT trial is in progress.

In conclusion, operator-dependent variability of size measurements of small nodules examined with LDTSCT is not negligible and should be considered when size thresholds are used in lung cancer screening.

REFERENCES

1. Yankelevitz DF, Reeves AP, Kostis WJ, et al. Small pulmonary nodules: volumetrically determined growth rates based on CT evaluation. *Radiology*. 2000;217:251–256.
2. Revel MP, Lefort C, Bissery A, et al. Pulmonary nodules: preliminary experience with three-dimensional evaluation. *Radiology*. 2004;231:459–466.
3. Ko JP, Rusinek H, Jacobs EL, et al. Small pulmonary nodules: volume measurement at chest CT – phantom study. *Radiology*. 2003;228:864–870.
4. Wormanns D, Kohl G, Klotz E, et al. Volumetric measurements of pulmonary nodules at multi-row detector CT: in vivo reproducibility. *Eur Radiol*. 2004;14:86–92.
5. Goodman LR, Gulsun M, Washington L, et al. Inherent variability of CT lung nodule measurements in vivo using semiautomated volumetric measurements. *AJR Am J Roentgenol*. 2006;186:989–994.
6. Marten K, Auer F, Schmidt S, et al. Inadequacy of manual measurements compared to automated CT volumetry in assessment of treatment response of pulmonary metastases using RECIST criteria. *Eur Radiol*. 2006;16:781–790.
7. Bolte H, Riedel C, Jahnke T, et al. Reproducibility of computer-aided volumetry of artificial small pulmonary nodules in ex vivo porcine lungs. *Invest Radiol*. 2006;41:28–35.
8. 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*. 2005;237:395–400.
9. Libby DM, Smith JP, Altorki NK, et al. Managing the small pulmonary nodule discovered by CT. *Chest*. 2004;125:1522–1529.
10. Henschke CI, McCauley DI, Yankelevitz D, et al. Early Lung Cancer Action Project: overall design and findings from baseline screening. *Lancet*. 1999;354:99–105.
11. Claudia I. Henschke. International Early Lung Cancer Action Program: Protocol. February 28, 2006. Available at: http://www.ielcap.org/ielcap.pdf. Accessed August 22, 2006.
12. Picozzi G, Paci E, Lopes Pegna A, et al. Screening of lung cancer with low dose spiral CT: results of a three year pilot study and design of the randomised controlled trial "Italung-CT." *Radiol Med*. 2005;109:17–26.
13. Kostis WJ, Reeves AP, Yankelevitz DF, et al. Three-dimensional segmentation and growth-rate estimation of small pulmonary nodules in helical CT images. *IEEE Trans Medical Imaging*. 2003;22:1259–1274.

14. British Standards Institution. *Precision of Test Methods, Part I: Guide for Determination of Repeatability and Reproducibility for a Standard Test Method*. London, England: BSI. BS 5497, part I; 1979.
15. Rousson V, Gasser T, Seifert B. Assessing intrarater, interrater and test-retest reliability of continuous measurements. *Stat Med*. 2002;21: 3431–3446.
16. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1:307–310.
17. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res*. 1999;8:135–160.
18. Halligan S. Reproducibility, repeatability, correlation and measurement error. *Br J Radiol*. 2002;75:193–194; discussion, 194–195.
19. World Health Organization. *WHO Handbook for Reporting Results of Cancer Treatment*. Geneva, Switzerland: World Health Organization; 1979.
20. Therasse P, Arbuck SG, Eisenhauer EA, 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*. 2000;92:205–216.
21. Prasad SR, Saini S, Sumner JE, et al. Radiological measurement of breast cancer metastases to lung and liver: comparison between WHO (bidimensional) and RECIST (unidimensional) guidelines. *J Computer Assist Tomogr*. 2003;27:380–384.
22. Revel MP, Bissery A, Bienvenu M, et al. Are two-dimensional CT measurements of small noncalcified pulmonary nodules reliable? *Radiology*. 2004;231:453–458.
23. Erasmus JJ, Gladish GW, Broemeling L, et al. Interobserver and intraobserver variability in measurements of non-small-cell carcinoma lung lesions: implications for assessment of tumor response. *J Clin Oncol*. 2003;21:2574–2582.
24. Kuhnigk JM, Dicken V, Bornemann L, et al. Morphological segmentation and partial volume analysis for volumetry of solid pulmonary lesions in thoracic CT scans. *IEEE Trans Med Imaging*. 2006;25:417–434.
25. Kostis WJ, Yankelevitz DF, Reeves AP, et al. Small pulmonary nodules: reproducibility of three-dimensional volumetric measurement and estimation of time to follow-up CT. *Radiology*. 2004;231:446–452.
26. Ko JP, Marcus R, Bomsztyk E, et al. Effect of blood vessels on measurement of nodule volume in a chest phantom. *Radiology*. 2006; 239:79–85.
27. Boll DT, Gilkeson RC, Fleiter TR, et al. Volumetric assessment of pulmonary nodules with ECG-gated MDCT. *AJR Am J Roentgenol*. 2004;183:1217–1223.
28. Pastorino U, Bellomi M, Landoni C, et al. Early lung-cancer detection with spiral CT and positron emission tomography in heavy smokers: 2 years results. *Lancet*. 2003;362:593–597.
29. Henschke CI, Yankelevitz DF, Naidich DP, et al. CT screening for lung cancer: suspiciousness of nodules according to size on baseline scans. *Radiology*. 2004;231:164–168.
30. Xu D, Vliegenthart R, Wang Y, et al. Baseline findings of a randomized Ducth-Belgian trial on lung cancer screening by low-dose CT (NELSON study). *Eur Radiol*. 2006;(Suppl 1):144.
31. Henschke CI, Yankelevitz DF, Mirtcheva R, et al. CT screening for lung cancer: frequency and significance of part-solid and nonsolid nodules. *AJR Am J Roentgenol*. 2002;178:1053–1057.
32. Reeves AP, Chan AB, Yankelevitz DF, et al. On measuring the change in size of pulmonary nodules. *IEEE Trans Medical Imaging*. 2006;25: 435–450.