Pulmonary Nodules: Interscan Variability of Semiautomated Volume Measurements with Multisection CT—Influence of Inspiration Level, Nodule Size, and Segmentation Performance[1]

Hester A. Gietema, MD, PhD
Cornelia M. Schaefer-Prokop, MD, PhD
Willem P. T. M. Mali, MD, PhD
Gerard Groenewegen, MD, PhD
Mathias Prokop, MD, PhD

Purpose: To prospectively assess the precision of semiautomated volume measurements of pulmonary nodules at low-dose multi–detector row computed tomography (CT) and to investigate the influence of nodule size, segmentation algorithm, and inspiration level.

Materials and Methods: This study had institutional review board approval; written informed consent was obtained from all patients. Between June 2004 and March 2005, 20 patients (15 men, five women; age range, 40–84 years; mean age, 57 years) referred for chest CT for known lung metastases underwent two additional low-dose chest CT examinations without contrast material (collimation, 16 × 0.75 mm). Between these examinations, patients got off and on the table to simulate the conditions for a follow-up examination. Noncalcified solid pulmonary nodules between 15 and 500 mm^3 that did not abut vessel or pleura were measured in both studies by using widely applied commercial semiautomated software. Interscan variability was established with the Bland and Altman approach. The impact of nodule shape (spherical or nonspherical) on measurement variability was assessed by using one-way analysis of variance, while the contributions of mean nodule volume and change in lung volume were investigated with univariate linear regression for completely (group A) and incompletely (group B) segmented nodules.

Results: Two hundred eighteen eligible nodules (volume range, 16.4–472.7 mm^3; 106 spherical, 112 nonspherical) were evaluated. The 95% confidence interval for difference in measured volumes was −21.2%, 23.8% (mean difference, 1.3%). The precision of nodule segmentation was highly dependent on nodule shape ($P < .001$) and was weakly related to inspiration level for completely segmented nodules ($r = -0.20$; $P < .047$), while mean nodule volume did not show any effect ($P = .15$ and $P = .81$ for group A and B nodules, respectively).

Conclusion: Variation of semiautomated volume measurements of pulmonary nodules can be substantial. Segmentation represents the most important factor contributing to measurement variability, while change in inspiration level has only a weak effect for completely segmented nodules.

[1] From the Departments of Radiology (H.A.G., W.P.T.M.M., M.P.) and Medical Oncology (G.G.), University Medical Center, Heidelberglaan 100, 3584 CX Utrecht, the Netherlands; and Department of Radiology, Academic Medical Center, Amsterdam, the Netherlands (C.M.S.). From the 2005 RSNA Annual Meeting. Received June 18, 2006; revision requested August 23; revision received October 30; accepted November 22; final version accepted April 3, 2007. **Address correspondence to** H.A.G. (e-mail: *h.gietema@umcutrecht.nl*).

Precise assessment of change in size of pulmonary nodules on follow-up scans is pivotal for evaluation of nodules in lung cancer screening trials (1–6) but also in clinical practice, where change in size is used to evaluate response to therapy (7). Today, in an oncologic setting, two-dimensional measurements are being performed by using electronic calipers to determine the longest diameter of the target lesion, as recommended by the Response Evaluation Criteria in Solid Tumors, or RECIST, group (7). However, three-dimensional measurements have been shown to be more accurate (4,5) and are therefore often applied in lung cancer screening trials (1,3). Moreover, RECIST criteria apply to nodules larger than 10 mm, while in lung cancer screening trials, only nodules smaller than 10 mm are followed up.

Volume doubling times of malignant pulmonary nodules may vary between 30 and 400 days (6). For small or moderately growing nodules, the increase in size over a follow-up period of up to 1 year will be only in the range of a few voxels (2,6). Therefore, it is crucial to distinguish between real but slow changes in size and other factors that influence volume measurements. Patient position, heart pulsation, and inspiration levels have been hypothesized to influence the assessment of nodule size (3). Interscan variability has already been shown to be an important factor by Wormanns et al (8), but only patient positioning was considered in that study. Kostis et al (3) evaluated the reproducibility of repeat volume measurements of stable small pulmonary nodules that showed no increase or decrease in size for more than 2 years. As a consequence of the inclusion criteria, the vast majority of nodules were smaller than 5 mm in diameter, and no distinction could be made between truly identical volumes over the follow-up period and minor changes in size during this period.

The purpose of our study was to prospectively assess the precision of semiautomated volume measurements of pulmonary nodules at low-dose multidetector row computed tomography (CT) and to investigate the influence of nodule size, segmentation algorithm, and inspiration level.

Advances in Knowledge

- The nodule segmentation algorithm is the main source of variability of volume measurements of pulmonary nodules for the tested software.
- Variability of volume measurements for completely segmented nodules on repeat scans was weakly correlated with variability in inspiration ($r = -0.20$, $P = .047$).

Materials and Methods

Patients and Nodule Selection

The study was approved by the institutional review board of University Medical Center, and written informed consent was obtained from all patients after the risks of the additional radiation dose were explained. Between June 2004 and March 2005, we enrolled 20 consecutive adult patients (15 men, five women; age range, 40–84 years; mean age, 57 years) who had known pulmonary metastases. All patients visited the oncology outpatient department and were referred for chest CT for clinical indications. The presence of lung metastases had been previously shown at chest CT or chest radiography. The majority of patients ($n = 13$) were referred for chest CT to monitor the effect of anticancer therapy. The remaining patients ($n = 7$) were referred for baseline chest CT before the start of anticancer therapy. These patients had pulmonary metastases at chest radiography. The underlying primary tumors were melanoma ($n = 3$), renal cell carcinoma ($n = 6$), colorectal cancer ($n = 5$), breast carcinoma ($n = 2$), prostate cancer ($n = 1$), seminoma ($n = 1$), medullar thyroid cancer ($n = 1$), and esophageal adenocarcinoma ($n = 1$).

Even though the patients also had larger lesions, we included only nodules smaller than 500 mm^3, a size that corresponds to a mean diameter of approximately 10 mm (exact value, 9.85 mm), because the commercially available algorithm we applied (LungCare; Siemens Medical Solutions, Erlangen, Germany) has been released for use with nodules smaller than 10 mm in diameter. The minimum nodule volume we included in this evaluation was 15 mm^3 (corresponding to a diameter of about 3 mm). We included not only nodules suspected of being metastases but also nodules that could potentially have benign histologic features. Completely calcified nodules, however, were excluded. We included only solid nodules in our evaluation because the current software is not released for use in volume measurements of nonsolid and part-solid nodules.

Image Acquisition

We performed two low-dose chest CT examinations without contrast material, followed by a contrast material–enhanced standard-dose chest CT examination for clinical purposes. Between the two low-dose examinations, patients were asked to get off and on the table to simulate the conditions of a repeat examination for follow-up of a pulmonary nodule. With that setup, growth or decrease in size of the lung lesions could reliably be excluded.

All scans were acquired with a 16-sec-

Implication for Patient Care

- Precise assessment of change in size of pulmonary nodules on follow-up scans is pivotal for evaluation of nodules in lung cancer screening trials but also in clinical practice, where change in size is used to evaluate response to therapy.

Published online before print
10.1148/radiol.2452061054

Radiology 2007; 245:888–894

Author contributions:
Guarantor of integrity of entire study, H.A.G.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; manuscript final version approval, all authors; literature research, H.A.G., M.P.; clinical studies, all authors; statistical analysis, all authors; and manuscript editing, all authors

Authors stated no financial relationship to disclose.

tion CT scanner (Mx8000 IDT; Philips Medical Systems, Cleveland, Ohio) by using a spiral mode with 16 × 0.75-mm collimation. The entire chest was scanned in about 10 seconds by using a caudocranial scan direction. Examinations were performed in full inspiration after appropriate instructions were given to the patients. So as to reproduce the standard situation in a screening setting, we used no spirometric control or respiratory belt. Exposure settings for the additional low-dose examinations were 30 mAs and 120 kVp (volume CT dose index [$CTDI_{vol}$] = 2.2 mGy) for patients who weighed 80 kg or less and 30 mAs and 140 kVp for those who weighed more than 80 kg ($CTDI_{vol}$ = 3.5 mGy). Transverse images were reconstructed at a thickness of 1.0 mm and a 0.7-mm increment by using a moderately soft reconstruction kernel (kernel B), the smallest field of view that included the outer rib margins at the widest dimension of the thorax, and a 512 × 512 matrix.

Semiautomated Volume Measurements of Pulmonary Nodules

Data were transferred from the CT scanner to a digital workstation (Leonardo; Siemens Medical Solutions) with commercially available software for semiautomated volume measurements (LungCare; Siemens Medical Solutions). Nodules were identified by a single observer (H.A.G., who had 1 year of experience in radiology and who had been trained for this specific task) by using transverse thin-slab maximum intensity projections (slab thickness, 10 mm) that were displayed with window width and level settings of 1500 and −500 HU, respectively. After a candidate nodule is manually marked with a mouse click in the center of the nodule, the program automatically defines a volume of interest around the nodule that can be analyzed by using volume-rendered displays. Quantitative evaluation of a nodule is initiated with a second mouse click, which starts an automated volume measurement program that has been described previously (8). In this step, the nodule of interest is segmented and the volume of the segmented area is calculated. This segmented area is shown by a yellow overlay on the nodule. No manual interaction was performed to correct mismatches.

To minimize the influence of the separation process, which distinguishes between the nodule itself and adjacent structures, we included only lesions that had no direct contact with the pleura or vessels. That way, it was ensured that we focused solely on the precision of measurement of the volume of an isolated solid lesion. For each patient, all nodules with a volume of 15–500 mm^3 that met the inclusion criteria were included for evaluation.

Nodules in the second study were identified with knowledge of findings on the first study; the nodules were matched by using the combination of section number, lung segment, and distance to the pleura.

Evaluation of Effect of Segmentation Performance

We visually determined the precision of the measurement software by assessing whether the nodule was completely segmented. Nodules were categorized into two groups (A and B) on the basis of whether the yellow overlay completely matched the nodule (group A) or whether visual assessment determined a mismatch (group B). Mismatch was defined as the visually obvious exclusion of a part of the investigated nodule from the segmented area.

Evaluation of Nodule Shape

In a separate reading session, an experienced observer (M.P., with 20 years of experience in radiology) visually assessed nodule shape and categorized nodules into three subgroups. A nodule was defined as spherical when it had a constant radius and as lobular when it had a variable radius but smooth outer margins. It was defined as irregular when the outer margins were not smooth.

To investigate the effect of nodule shape on the performance of the segmentation algorithm, we determined how many nodules were spherical, nonspherical, or irregular for completely (group A) or incompletely (group B) segmented nodules.

Evaluation of Inspiration Level and Lung Volume

To compare the inspiration levels of the two scans, we calculated the lung volume for every scan by using completely automated software developed in house. The algorithm is similar to one described previously (9). The lungs are segmented from adjacent soft-tissue structures (eg, the mediastinum, vascular structures, and the chest wall) by region-growing from an automated starting point in the trachea, including all connected areas with attenuation of less than −500 HU. In a second step, the trachea and main bronchi are excluded from the lungs. The number of voxels within the segmented lungs is multiplied by voxel size to calculate total lung volume.

Natural variation in inspiration level between the two low-dose scans was established as the ratio between the lung volume on the second scan and the lung volume on the first scan.

Statistical Evaluation

Nodule volumes are given as means ± standard deviations. Reproducibility of volume measurements was assessed by correlating nodule volumes on both scans by using the Spearman correlation coefficient for non–normally distributed populations.

Differences in volume (ΔV) were calculated by subtracting the volume measured on the first scan (V_1) from the volume measured on the second scan (V_2). These differences were plotted against the mean nodule volume by using the approach described by Bland and Altman (10):

$$\Delta V = \frac{V_2 - V_1}{(V_2 + V_1)/2}.$$

The differences in volume measurements were normalized with respect to mean nodule volume to assess relative differences with the following formula:

$$\Delta V_{rel} = 100\% \cdot \frac{V_2 - V_1}{(V_1 + V_2)/2}.$$

Limits of agreement were given as 95% confidence intervals. Interscan variability was defined as the 95% confidence interval of the relative differences. An increase in nodule volume above these upper limits of agreement can, with 95% confidence, be attributed to real growth.

Because the shape of a nodule was a binary variable, an analysis of variance test to compare the variances of both groups was performed to assess the effect of nodule shape on the relative difference in measured volumes. To assess the effects of inspiration level, we performed univariate regression for both groups, with the relative difference as the dependent variable and the ratio of lung volumes as the independent variable. To assess the impact of nodule volume on measurement variation for both groups, we performed univariate linear regression with the coefficient of variation (standard deviation divided by mean volume) of both measurements as the dependent variable and the mean nodule volume as the independent variable. All statistics were calculated with software (SPSS, version 12.0; SPSS, Chicago, Ill). *P* values less than .05 were considered to indicate significant differences.

Figure 1

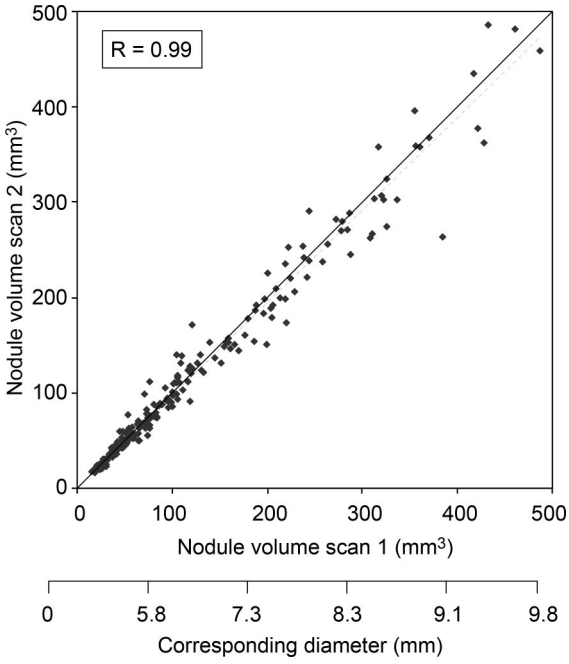

Figure 1: Graph shows volumetric results for each nodule. Dashed line shows correlation between both measurements of same nodule. Note that this line is almost equal to the (continuous) $x = y$ line.

Results

Nodule Characteristics

A total of 218 noncalcified solid intraparenchymal nodules with a volume between 15 and 500 mm^3 were eligible for analysis. Twelve calcified nodules and

Figure 2

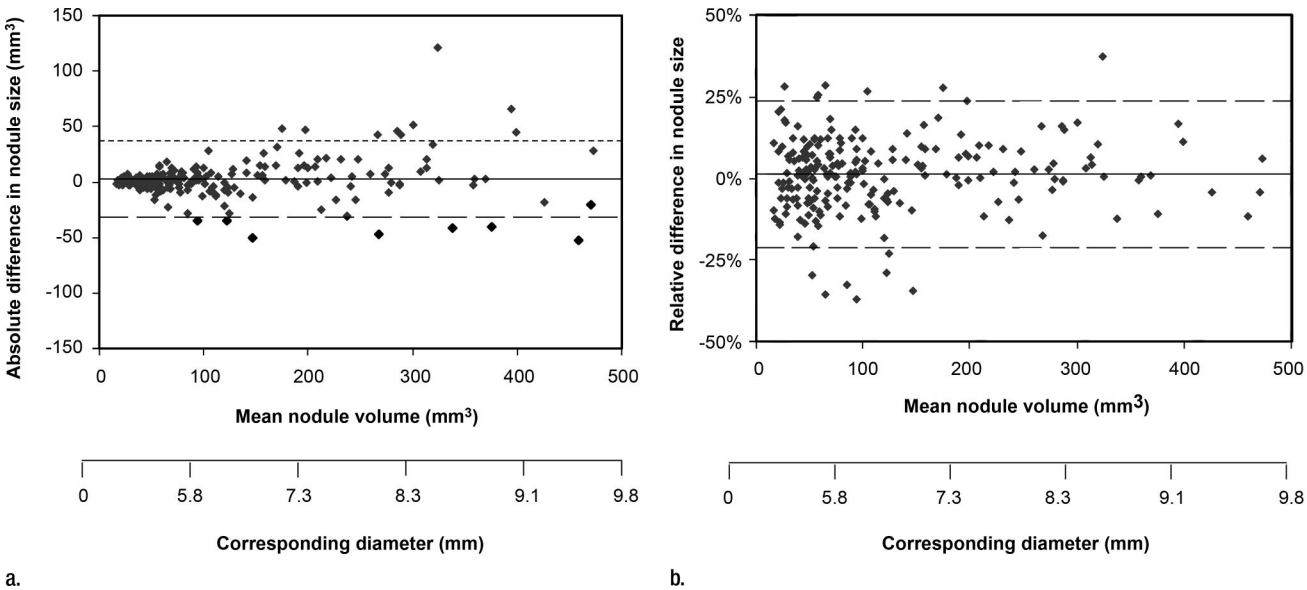

Figure 2: Graphs show interscan agreement of volume measurements of all nodules. **(a)** Absolute and **(b)** relative differences between both measurements are plotted against mean nodule volume. Mean difference is shown by continuous line; upper and lower limits of agreement are shown by dashed lines.

Influence of Nodule Shape on Absolute and Relative Difference in Volume

Group	No. of Nodules	Size (mm^3)*	Change in Volume (mm^3)*	Relative Difference (%)†	Positive Difference (%)‡
A: Completely segmented	106	16.4–369.2	0.04–15.35	0.28 ± 6.2	4.2 (0.0–34)
B: Incompletely segmented	112	21.4–472.7	2.5–120.79	1.61 ± 14.5	12.4 (1.0–37)

* Data are ranges.

† Data are mean values ± standard deviations.

‡ Data are mean values, with ranges in parentheses.

12 pleura-based nodules were excluded from analysis. No vessel-attached nodules were detected.

The mean volume of all nodules was 123.0 mm^3 ± 101.9 (range, 16.4–472.7 mm^3; median, 82.7 mm^3). The number of eligible nodules per patient ranged from zero (no metastases visible after therapy) in four patients to 62 nodules in one patient. The median number of nodules per patient was six. None of the patients had only nodules that were larger than 500 mm^3.

One hundred six nodules were spherical, 30 nodules had a nonspherical shape, and 82 nodules showed irregular margins.

Reproducibility of Volume Measurements

The reproducibility of volume measurements for the whole group was excellent, with a Spearman correlation coefficient of 0.99 (Fig 1). For the total group of nodules, the mean difference in volume measurements amounted to 2.4 mm^3, ranging from −53.0 to 120.8 mm^3 (95% confidence interval: −32.0, 36.7 mm^3) (Fig 2a). The mean relative difference amounted to 1.3%, with a 95% confidence interval of −21.2%, 23.8% (Fig 2b).

Influence of Nodule Size, Segmentation Performance, and Inspiration Level

Analysis of segmentation performance revealed that 106 of the 218 nodules (48.6%) had been completely segmented (group A: nodule size range, 16.4–369.2 mm^3), while 112 nodules (51.4%) were incompletely segmented (group B: nodule size range, 21.4–472.7 mm^3; Table; Fig 3). All nodules were categorized into the same subgroup for both scans.

The group of completely segmented nodules (group A) contained only spherical lesions. All nodules with irregular margins (n = 82) or a nonspherical shape (n = 30) were incompletely segmented (group B). Group B did not contain any spherical nodules.

Interscan variability was −11.9% to 12.4% for completely segmented nodules. For incompletely segmented nodules, this interval was more than twice as large: −26.8% to 30.0%. The maximum absolute difference between measured volumes on consecutive scans was 15.35 mm^3 for completely segmented nodules and 120.79 mm^3 for incompletely segmented ones. The maximum relative difference between volumes was 15% for completely segmented and 37% for incompletely segmented nodules. The measurement variability, as given by the width of the 95% confidence interval standard deviations, was 2.3 times smaller for completely segmented nodules than for incompletely segmented ones. This difference was statistically significant ($P < .002$, F test).

Figure 3

a. b.

Figure 3: **(a)** Completely and **(b)** incompletely segmented nodule. Note that yellow overlay does not cover whole nodule in **b**. Boxes in lower right corners show three-dimensional orientation of images.

Ratio of lung volume between the second and the first scan ranged from 88% to 116% (Fig 4). Lung volume calculation failed in five patients because of a high number of nodules (n = 2) or fibrosis (n = 3), while lung volumes of the four patients without nodules could not be used for analysis.

The coefficient of variation in nodule volume was not a significant predictor of relative difference in nodule volume for either completely (P = .15) or incompletely (P = .81) segmented nodules.

For completely segmented nodules, the relative difference decreased with increasing ratio of inspiration level (r = −0.20, P = .047), while for incompletely segmented nodules, we found no correlation (P = .67).

Discussion

In this study we tested the precision of a commercially available algorithm for the assessment of nodule volumetry. To

minimize effects of segmentation from adjacent structures, only nodules completely surrounded by lung tissue were included in the study group. Results of previous studies have shown that the applied software is very accurate for small spherical nodules (11), but our results suggest that the precision may vary substantially with nodule morphology. While the precision was extremely high for spherical solid nodules, the precision decreased for nonspherical nodules and nodules with irregular shapes. This is especially noteworthy considering the fact that many nodules detected in a lung cancer screening setting do not have perfectly spherical shapes with smooth margins. Thus, assessment of the volume of such nodules with the semiautomated software we tested might be prone to considerable variation.

Wormanns et al (8) applied a study design comparable to ours and used the same type of semiautomated software. Although we used thinner collimation and included a higher number of nodules, we found comparable limits of agreement: −21.2% to 23.8% in our study versus −20.6% to 21.9% in the study by Wormanns et al. Wormanns and colleagues discussed that both ill-defined shape and attachment to pleura or vessels of some of the included nodules could explain why reported variability was higher in vivo than what has been seen in phantom studies. However, they did not further analyze the contribution of these factors separately. We specifically excluded pleura-based and vessel-attached nodules to be able to assess these factors separately. The reason we still found similar limits of agreement despite these exclusion criteria is likely to be due to the higher number of irregularly shaped nodules in our study than in the group of nodules analyzed by Wormanns et al.

Although LungCare uses a global thresholding method for the segmentation of a candidate nodule, the algorithm reported by Kostis et al (12) uses a more sophisticated segmentation approach that was developed to deal with irregular shapes. With their algorithm, Kostis et al also found an effect of the initial nodule size on the extent of variability, while Wormanns et al had reported similar limits of agreement for different size ranges of nodules. We could not prove a significant effect of lesion size on the variability of measurements for nodules that had been perfectly segmented. One could argue that the precision of the segmentation algorithm could have been improved by manual modification. Our software offers the possibility of adapting the segmented area by modifying the cutoff value of the cross-section curves. The means of that adaptation process, however, are limited with respect to the fact that modification of the cutoff value is only valid for the complete circumference of the nodule. As a consequence, moving the cutoff value of the cross-section curve to the right (resulting in increased overlaid volume) helps to include parts of the nodules that were originally not included, but at the expense of including surrounding parts, leading to an overestimation of the volume. Other important aspects are that the use of any manual adaptation or modification of the volumetric measurement would not only increase the radiologist's reading time but would also introduce another variable.

Changes in inspiratory level turned out to be another source of variability. We found that higher inspiration level led to a decrease in measured volume. This finding is most likely due to the fact that the attenuation of the surrounding lung parenchyma changes with inspiratory level, leading to an alteration of the cutoff value and thus the volume assessment. Although the contribution of inspiratory level was statistically significant, the quantities of volume changes introduced by it were only small.

Other potential sources of variability are interobserver and intraobserver variability. Because in our study, the same single observer measured all nodules, interobserver variability did not contribute to the variability of our measurements. We did not specifically evaluate intraobserver variability, but Wormanns and colleagues have already shown that the effect of intraobserver variability is negligible compared with the effect of interscan variability (8).

The main limitation of our study was that all results reported are valid only for the particular software release we used. Although this program is widely used, our results are not transferable to other algorithms. The limitation we found is related to the fact that a global thresholding method is used for the segmentation of a candidate nodule. Algorithms that use a more sophisticated segmentation algorithm are likely to also achieve a higher precision for irregularly shaped nodules.

In conclusion, the precision of the commercially available software tool tested was dependent on lung nodule morphology and was found to be less reproducible for nonspherical than for spherical solid nodules. The extent of variability decreased with increasing nodule size and higher inspiration level.

Taking the reported variation into

Figure 4

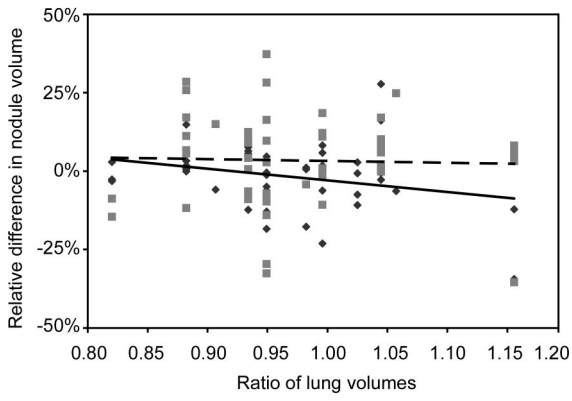

Figure 4: Graph shows correlation between ratios of lung volumes (lung volume on second scan divided by lung volume on first scan) and differences in nodule size. Black diamonds show completely segmented nodules; gray squares show incompletely segmented nodules. For incompletely segmented nodules, no correlation could be demonstrated, while for completely segmented nodules, nodule volume decreased moderately but significantly with increase in lung volume ($r = 0.20$, $P < .001$).

account, the threshold for calling an increased measured volume of a lung nodule a real volume increase with 95% confidence lies at a 30% increase for an irregularly shaped lesion. For spherical nodules, this threshold can be lowered to 15%.

References

1. Gietema HA, Wang Y, Xu D, et al. Pulmonary nodules detected by lung cancer screening: inter-observer variability of semi-automated volume measurements. Radiology 2006;241(1):251–257.
2. Hasegawa M, Sone S, Takashima S, et al. Growth rate of small lung cancers detected on mass CT screening. Br J Radiol 2000; 73(876):1252–1259.
3. Kostis WJ, Yankelevitz DF, Reeves AP, Fluture SC, Henschke CI. Small pulmonary nodules: reproducibility of three-dimensional volumetric measurement and estimation of time to follow-up CT. Radiology 2004; 231(2):446–452.
4. Revel MP, Lefort C, Bissery A, et al. Pulmonary nodules: preliminary experience with three-dimensional evaluation. Radiology 2004; 231(2):459–466.
5. Revel MP, Bissery A, Bienvenu M, Aycard L, Lefort C, Frija G. Are two-dimensional CT measurements of small noncalcified pulmonary nodules reliable. Radiology 2004; 231(2):453–458.
6. Usuda K, Saito Y, Sagawa M, et al. Tumor doubling time and prognostic assessment of patients with primary lung cancer. Cancer 1994;74(8):2239–2244.
7. Therasse P, Arbuck SG, Eisenhauer EA, et al. New guidelines to evaluate the response to treatment in solid tumors. J Natl Cancer Inst 2000;92(3):205–216.
8. 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(1):86–92.
9. Hu S, Hoffman EA, Reinhardt JM. Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. IEEE Trans Med Imaging 2001;20(6):490–498.
10. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1(8476): 307–310.
11. Wormanns D, Fiebich M, Saidi M, Diederich S, Heindel W. Automatic detection of pulmonary nodules at spiral CT: clinical application of a computer-aided diagnosis system. Eur Radiol 2002;12(5):1052–1057.
12. Kostis WJ, Reeves AP, Yankelevitz DF, Henschke CI. Three-dimensional segmentation and growth-rate estimation of small pulmonary nodules in helical CT images. IEEE Trans Med Imaging 2003;22(10):1259–1274.