Chapter 6

Interscan variability of semiautomated volume measurements in intraparenchymal pulmonary nodules using multidetector-row computed tomography: Influence of inspirational level, nodule size and segmentation performance

Hester Gietema
Cornelia Schaefer-Prokop
Willem Mali
Gerard Groenewegen
Mathias Prokop

ABSTRACT

PURPOSE

To prospectively assess the reproducibility of semiautomated volume measurements of pulmonary nodules in low-dose multidetector-row computed tomography (CT) and to investigate the influence of nodule size, segmentation algorithm and inspirational level.

MATERIAL AND METHODS

The study was approved by our institutional review board and written informed consent was obtained from all patients. Between June 2004 and March 2005, 20 patients (15 men, 5 women, aged 40-84yrs, mean 57yrs) referred for chest CT for known lung metastases, received two additional low-dose non-contrast-enhanced chest CTs (16 x 0.75mm collimation). Between these scans, patients got off and on the table to simulate a follow-up scan. Non-calcified solid pulmonary nodules between 15 and 500 mm^3, not abutting with vessel or pleura, were measured on both scans using widely applied commercial semiautomated software. Inter-scan variability was established by the Bland-and-Altman approach. We assessed the impact of nodule shape (spherical or non-spherical) to measurement variability using one-way ANOVA, while the contributions of mean nodule volume and change in lung volume were investigated with univariate linear regression for completely (group A) and incompletely segmented (group B) nodules.

RESULTS

We evaluated 218 eligible nodules (range 16.4 to 472.7 mm^3, 106 spherical, 112 non-spherical shape). The 95% confidence interval for difference in measured volumes was -21.2% to 23.8% (mean difference 1.3%). The preciseness of nodule segmentation was highly dependent on nodule shape (p<0.001) and weakly related to inspirational level for completely segmented nodules (r=-0.20; p<0.047), while mean nodule volume did not show any impact (p=0.15 and p=0.81 for group A&B 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 inspirational level has only a weak impact for completely segmented nodules.

INTRODUCTION

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 oncological setting two-dimensional measurements are being performed using electronic calipers, in order to determine the longest diameter of the target lesion, as recommended by the 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 over 10mm, while in a lung cancer screening trials, only nodules less than 10mm are being 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 one year will be only in the range of a few voxels [2;6]. Therefore, it is crucial to distinguish between real but slow change in size and other factors influencing volume measurements. Patient position, heart pulsation and inspirational levels have been hypothesized to influence the assessment of nodule size [3]. Inter-scan variability has already been shown to be an important factor by Wormanns et al, but solely patient positioning was considered in that study [8]. Kostis et al evaluated the reproducibility of repeat volume measurements of stable small pulmonary nodules that showed no increase or decrease in the size of the nodule for more than 2 years. As a consequence of the inclusion criteria, the vast majority of nodules was smaller than 5mm in diameter and no distinction could be made between truly identical volume over the follow-up period or minor changes in size during this period [3].

Objective of our study was to evaluate the preciseness of a commercially available and widely used semiautomated volume measurement algorithm (LungCare©, Siemens Medical Solutions, Erlangen, Germany) for intrapulmonary nodules with a diameter of 3 mm to 10 mm (corresponding to a volume range of 15 mm[3] to 500 mm[3]). For this purpose, we quantified variation of semiautomated volume measurements of intrapulmonary nodules in repeat scans. Moreover, we investigated the contributions of nodule size, the performance of the segmentation algorithm and of variation in inspirational level to the extent of interscan variability.

MATERIALS AND METHODS

PATIENTS AND NODULE SELECTION

The study was approved by our institutional review board and written informed consent was obtained from all patients after explanation about the risks of the additional radiation dose. Between June 2004 and March 2005, we enrolled 20 consecutive adult patients (15 men, 5 women, 40-84 years old, mean 57 yrs) in this study who had known pulmonary metastases. All patients visited the oncology outpatient department and were referred for a chest CT for clinical indications. The presence of lung metastases had been previously shown on chest CT or chest radiography. The majority of patients (n =13) were referred for chest CT to monitor the effect of anticancer therapy. The other patients (n=7) were referred for a baseline chest CT before the start of anticancer therapy. These patients had pulmonary metastases shown on chest X-ray. 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 only included nodules smaller than 500 mm³, a size that corresponds to a mean diameter of approximately 10 mm (exact number: 9.85 mm), because the commercially available algorithm we applied (LungCare©, Siemens Medical Solutions, Erlangen, Germany) has been released for nodules less than 10mm in diameter. The minimum nodule volume we included in this evaluation was 15mm³ (corresponding to a diameter of about 3 mm). We included not only nodules suspected for being metastases but also nodules that could potentially have a benign histology. Completely calcified nodules, however, were excluded. We included only solid nodules in our evaluation since the current software is not released for volume measurements of non-solid and part-solid nodules.

IMAGE ACQUISITION

We performed two low-dose non-contrast-enhanced chest CTs, followed by a contrast-enhanced standard-dose chest CT for clinical purposes. Between the two low-dose scans patients were asked to get off and on the table to simulate the conditions of a repeat scan for follow-up of a pulmonary nodule. Using that set up, growth or decrease in size of the lung lesions could reliably be excluded.

All scans were acquired on a 16-slice CT scanner (Mx8000 IDT, Philips Medical Systems, Cleveland, OH) using a spiral mode with 16 x 0.75mm collimation. The entire chest was scanned in about 10 seconds using a caudo-cranial scan direction. Scans were performed in full inspiration after appropriate instruction of the patients. In order 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 30mAs and 120kVp (volume CT dose

index, $CTDI_{vol}$=2.2mGy) for patients weighing ≤80kg, and 30mAs at 140kVp for those weighing over 80kg ($CTDI_{vol}$=3.5mGy). Axial images were reconstructed at 1.0 mm thickness and 0.7mm increment, using a moderately soft reconstruction kernel (kernel "B"), the smallest field of view (FOV) that included the outer rib margins at the widest dimension of the thorax and a 512x512 matrix.

SEMIAUTOMATED VOLUME MEASUREMENTS OF PULMONARY NODULES

Data were transferred from the CT scanner to a digital workstation (Leonardo Workstation, Siemens Medical Solutions, Erlangen, Germany) with commercially available software for semiautomated volume measurements (LungCare©, Siemens Medical Solutions Erlangen, Germany). Nodules were identified by a single observer (one year of experience in radiology and trained for this specific task) using axial thin slab maximum intensity projections (slab thickness 10mm) displayed with window/center settings of 1500/-500HU. After manually marking a candidate nodule with a mouse click in the center of the nodule, the program automatically defines a volume of interest (VOI) around the nodule, which can be analyzed using volume rendered displays. Quantitative evaluation of a nodule is initiated with a second mouse click starting an automated volume measurement program 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 carried out to correct mismatches.

To minimize the influence of the separation process that distinguishes between the nodule itself and adjacent structures we only included lesions that had no direct contact with the pleura or vessels. That way it was assured that we solely focused on precisenessthe volume of an isolated solid lesion. For each patient, all nodules with a volume of 15-500 mm³ meeting the inclusion criteria were included for evaluation.

Nodules on the second scan were identified with knowledge of the first scan matching the nodules by the combination of slice number, lung segment and distance to the pleura.

EVALUATION OF THE EFFECT OF SEGMENTATION PERFORMANCE

We visually determined the preciseness of the measurement software by assessing if the nodule was completely segmented. Nodules were categorized in two groups (A and B) based on whether the yellow overlay completely matched the nodule (A) or whether visual assessments determined a mismatch (B). Mismatch was defined as a 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 (MP, 20 years of experience in radiology) visually assessed nodule shape and nodules were categorized into three subgroups. A nodule was defined as spherical when it showed a constant radius and as lobular when it showed a variable radius but with smooth outer margins. It was defined as irregular when the outer margins of the nodule were not smooth.

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

EVALUATION OF INSPIRATIONAL LEVEL AND LUNG VOLUME

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

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

STATISTICAL EVALUATION

Nodule volumes are given as mean ± standard deviation. Reproducibility of volume measurements was assessed by correlating nodule volumes in both scans using the Spearman's correlation coefficient for non-normally distributed populations.

Differences in volume (ΔV) were calculated by subtracting the volume measured on the

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

scan first (V_1) from the volume measured on the second scan (V_2). These differences were plotted against the mean nodule volume, using the approach described by Bland and Altman [10]:

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

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

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

Since the shape of a nodule was a binary variable, an ANOVA-test, comparing the variances of both groups, was performed to assess the impact of nodule shape on the relative difference in measured volumes. To assess the effects of inspirational level, we performed univariate regression for both groups with the relative difference as dependent variable and the ratio of lung volumes as independent variable. To assess the impact of nodule volume on measurement variation for both groups, we performed univariate linear regression with the relative SD (SD divided by mean volume) of both measurements as dependent variable and the mean nodule volume as independent variable. All statistics were calculated with SPSS statistical software package version 12.0 (SPSS, Chicago, USA). P-values <0.05 were considered significant.

RESULTS

NODULE CHARACTERISTICS

A total of 218 non-calcified, solid intraparenchymal nodules with a volume between 15 mm^3 and 500 mm^3 were eligible for analysis. Twelve calcified nodules and twelve pleura-based nodules were excluded from analysis. No vessel-attached nodules were detected.

Mean volume of all nodules was 123.0 mm^3 (±101.9 mm^3; range 16.4 mm^3 to 473.0 mm^3, median 82.7 mm^3). The number of eligible nodules per patient ranged from 0 (no metastases visible after therapy) in four patients to 62 nodules in one patient. The median number of nodules per patient was 6. None of the patients had only nodules that were larger than 500 mm^3.

Hundred six nodules were spherical, 30 nodules showed a non-spherical shape and 82 nodules showed irregular margins.
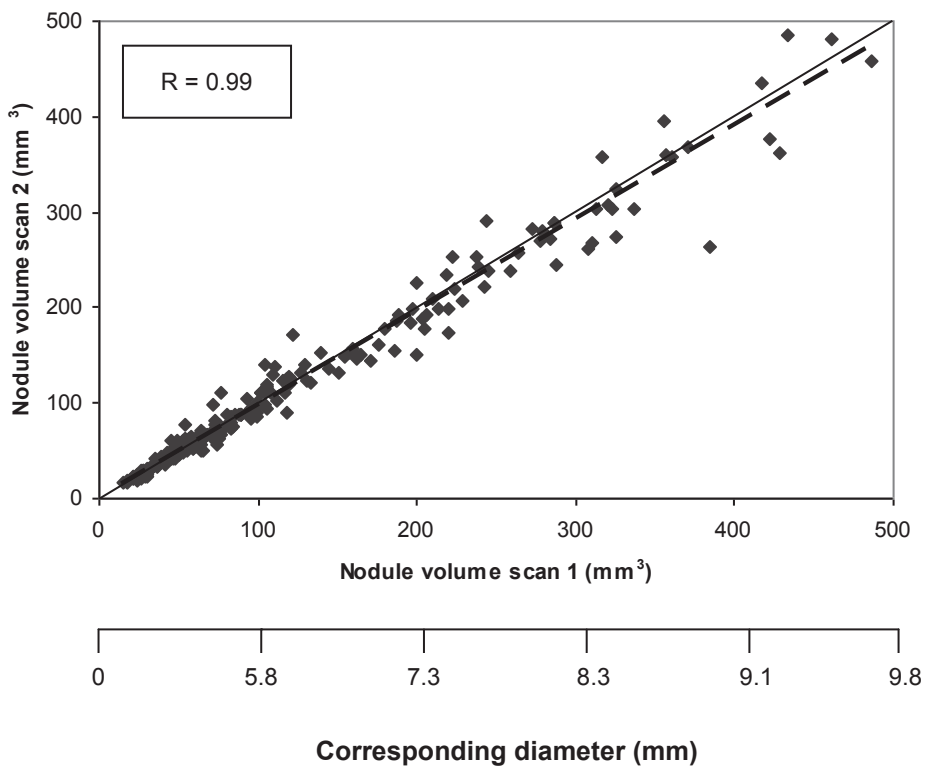
Figure 1

Volumetric results of each nodule. On the x-axis the volume measured on the first scan, on the y-axis the volume measured on the second scan. The dashed line shows the correlation between both measurements of the same nodule. Note that this line is almost equal to the (continuous) x=y line. Spearman's correlation coefficient = 0.99.

REPRODUCIBILITY OF VOLUME MEASUREMENTS

The reproducibility of volume measurements for the whole group was excellent with a Spearman's correlation coefficient of 0.99 (Figure 1). For the total group of nodules the mean difference in volume measurements amounted to $2.4mm^3$, ranging from $-53.0mm^3$ to $120.8mm^3$ (95% CI $-32.0mm^3$ to $36.7mm^3$, Figure 2A). The mean relative difference amounted to 1.3% with a 95% CI of -21.2% to 23.8% and a range of -37.1% to 37.3% (Figure 2B).

Group	Number of nodules	Size (mm3)	Change in volume (mm3)	Mean relative difference (±SD)	Mean positive difference (range)
Completely segmented	106	16.4 – 369.2	0.04 – 15.35	0.28% (±6.2%)	4.2% 0.0%-34%
Incompletely segmented	112	21.4 - 472.7	2.5 – 120.79	1.61% (±14.5%)	12.4% 1.0%-37%

Table 1
Influence of algorithm performance on absolute and relative difference in volume

2A

Corresponding diameter (mm)
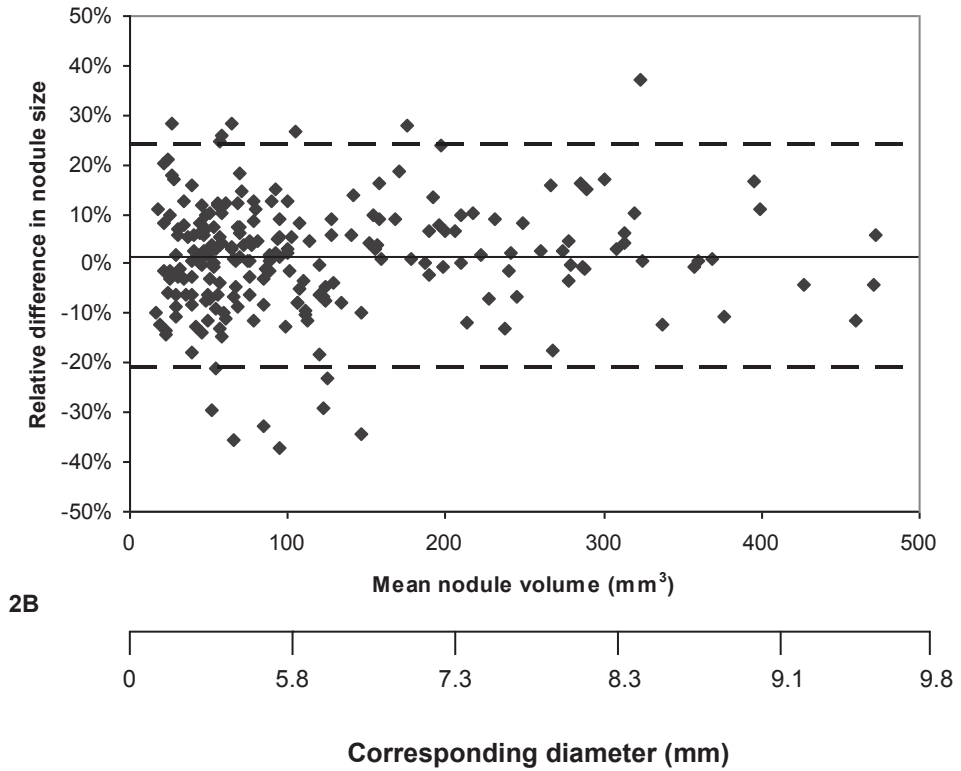
2B

Corresponding diameter (mm)
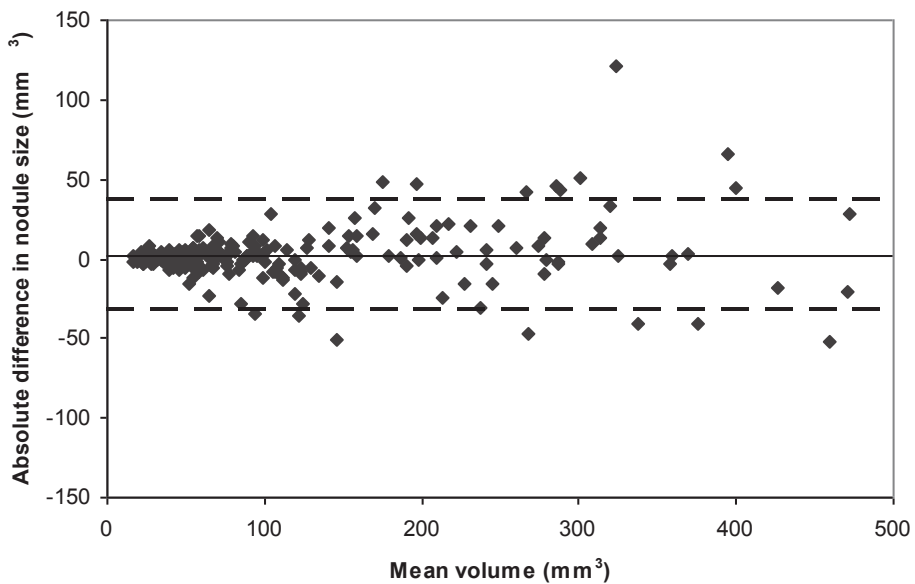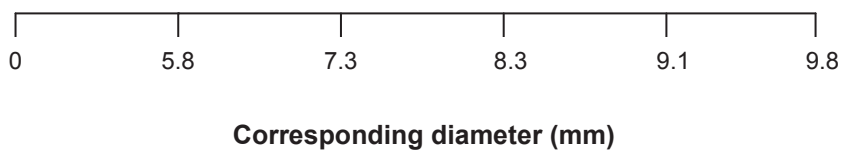
Figure 2
Inter-scan agreement of volume measurements of all nodules. The absolute (A) and relative (B) differences between both measurements are plotted against mean nodule volume. Mean difference is shown by the continuous line; upper and lower limits of agreement are shown by the dashed lines.
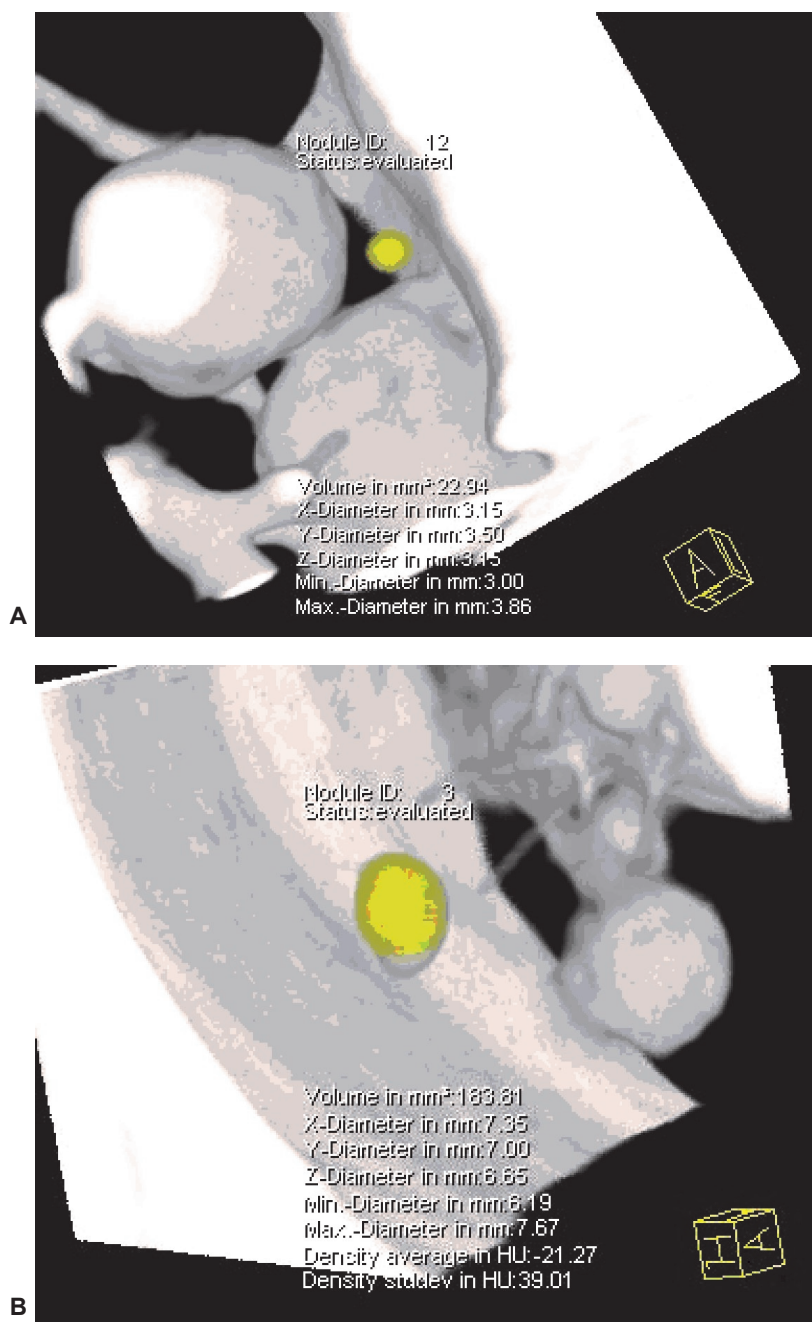
Figure 3

A complete (A) and an incomplete segmented nodule (B). Note that the yellow overlay does not cover the whole nodule in case of incomplete segmentation.

INFLUENCE OF NODULE SIZE, SEGMENTATION PERFORMANCE AND INSPIRATIONAL LEVEL

Analysis of the segmentation performance revealed that 106 of the 218 nodules (48.6%) had been completely segmented (group A, nodule size, $16.4mm^3$ to $369mm^3$) while 112 nodules (51.4%) were incompletely segmented (group B, nodule size, $21.4mm^3 - 473mm^3$ (Table 1). All nodules were categorized to the same subgroup on both scans.

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

Inter-scan 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 $15.4mm^3$ for completely segmented nodules and $121mm^3$ 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 (F-test, p< 0.0015).

Ratio in lung volume between the second and the first scan ranged from 88% to 116% (Figure 4). Lung volume calculation failed in five patients due to the 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.

No significant impact of mean nodule volume on the relative standard deviation was demonstrated for complete segmented nodules (p=0.15) or incomplete segmented nodules (p=0.81).

For complete segmented nodules, the relative difference decreased with increasing ratio of inspirational level (r=-0.20;p=0.047), while for incomplete segmented nodules we found no correlation (p=0.67).
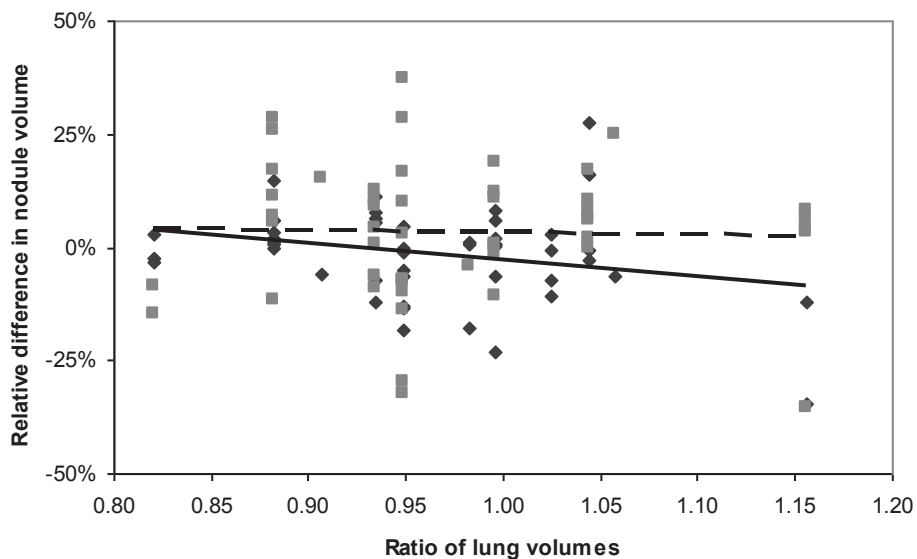
Figure 4

Correlation between ratios of lung volumes (lung volume on 2nd scan divided by lung volume on the 1st scan) and differences in nodule size. Black diamonds show completely segmented nodules, grey squares show incompletely segmented nodules. For incompletely segmented nodules, no correlation could be demonstrated (dashed line), while for completely segmented nodules, nodule volume decreased moderately, but significantly with increase in lung volume (r=0.20, p<0.001) as shown by the continuous line.

DISCUSSION

In this study we tested the preciseness of a commercially available and widely used 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. Previous studies have shown that the applied software was very accurate for small spherical nodules [11], but our results suggest that the preciseness may substantially vary with nodule morphology. While the preciseness was extremely high for spherical solid nodules, the preciseness went down for non-spherical 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 show perfectly spherical shape with smooth margins. Thus their volume assessment using the semiautomated software we tested might be prone to considerable variations.

Wormanns et al applied a study design comparable to ours, using the same type of semiautomated software [8]. Although we used thinner collimations 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. 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 seen in phantom studies. However, they did not further analyze the contribution of these factors separately. We specifically excluded pleural-based and vessel-attached nodules to be able to assess these factors separately. The reason why 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 as compared to the group of nodules analyzed by Wormanns et al.

While Lungcare© uses a global thresholding method for the segmentation of a candidate nodule, the algorithm reported by Kostis et al uses a more sophisticated segmentation approach, which is developed to deal with irregular shapes [12]. With their algorithm, Kostis et al also found an effect of the initial nodule size on the extent of variability, while Wormanns had reported similar limits of agreement for different size ranges of nodules. We could not show a significant effect of lesion size on the variability of measurements for nodules that had been perfectly segmented.

One could argue that the preciseness of the segmentation algorithm could have been improved by manual modification. Lungcare© offers the possibility to adapt the segmented area by modifying the cut-off value of the crossection curves. The means of that adaptation process, however, are limited with respect to the fact that modification of the cut-off value is valid for the complete circumference of the nodule. As a consequence, moving the cut-off value of the crossection curve to the right (resulting in increased overlaid volume) helps to include parts of the nodules 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 of the inspiratory level turned out to be another source of variability. We found that higher inspirational 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 cut-off 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 intra-observer variability. Since 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 intra-observer variability but Wormanns and colleagues had already shown that the extent of intraobserver variability was negligible compared to the effect of interscan variability [8].

The main limitation of our study is that all results reported are valid only for that particular software release. 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 achieve a higher preciseness also for irregularly shaped nodules.

CONCLUSION

The preciseness of the commercially available software tool tested is dependent on nodule morphology and was found to be less reproducible for non-spherical than for spherical solid nodules. The extent of variability decreases with increasing nodule size and higher inspirational level.

Taking the reported variation into account, the threshold to call an increased measured volume a real volume increase with a 95% confidence lies at 30% increase for an irregular shaped lesion. For spherical nodules this threshold can be lowered to 15%.