Effect of Nodule Characteristics on Variability of Semiautomated Volume Measurements in Pulmonary Nodules Detected in a Lung Cancer Screening Program

Ying Wang, MD
Rob J. van Klaveren, MD, PhD
Hester J. van der Zaag–Loonen, MD, PhD
Geertuida H. de Bock, PhD
Hester A. Gietema, MD, PhD
Dong Ming Xu, MD, PhD
Anne L. M. Leusveld, MD
Harry J. de Koning, MD, PhD
Ernst T. Scholten, MD
Johny Verschakelen, MD, PhD
Mathias Prokop, MD, PhD
Matthijs Oudkerk, MD, PhD

Purpose: To retrospectively assess volume measurement variability in solid pulmonary nodules (volume, 15–500 mm$^3$) detected at lung cancer screening and to quantify the independent effects of nodule morphology, size, and location.

Materials and Methods: This retrospective study was a substudy of the screening program that was approved by the Dutch Ministry of Health, and all participants provided written informed consent. Two independent readers used semiautomated software to measure the volume of pulmonary nodules detected in 6774 participants aged 50–75 years (5917 men). Nodules were classified according to their location (purely intraparenchymal, pleural based, juxtavascular, or fissure attached), morphology (smooth, polylobulated, spiculated, or irregular), and size ($\leq 50$ mm$^3$ or $>50$ mm$^3$). The level of agreement was expressed by using the absolute values of the relative volume differences (RVDs). Multivariate logistic regression analysis was performed, and odds ratios (ORs) were computed to quantify the independent effects of morphology, location, and size on RVD categories.

Results: Altogether, 4225 nodules in 2239 participants were included. Complete agreement in volume was obtained for 3646 (86%) of the nodules. Disagreement was small (absolute value of RVD $\leq 5\%$) for 173 (4%) nodules, moderate (absolute value of RVD $5\%$ but $<15\%$) for 232 (6%), and large (absolute value of RVD $\geq 15\%$) for 174 (4%). Multivariate analysis showed that the ORs of volume disagreement were 15.7, 3.1, and 1.9 for irregular, spiculated, and polylobulated nodules, respectively; 3.5, 2.6, and 2.1 for juxtavascular, pleural-based, and fissure-attached nodules, respectively; and 1.3 for large nodules compared with smooth, purely intraparenchymal, and small reference nodules.

Conclusion: Nodule morphology, location, and size influence volume measurement variability, particularly for juxtavascular and irregular nodules.

© RSNA, 2008
With the introduction of multidetector computed tomography (CT) and the ongoing advances in multidetector CT technology, the detection of small pulmonary nodules has become common, with detection rates ranging from 23% to 74% in lung cancer CT screening studies (1–5). Today, the challenge is to identify the malignant nodules among the majority of benign nodules. Volume doubling time has received attention as a diagnostic tool with which to differentiate malignant nodules from benign nodules, especially subcentimeter lesions in which additional diagnostic techniques—such as positron emission tomography, contrast material–enhanced CT, and percutaneous needle biopsy—are less suitable (6–8). Malignant solid pulmonary nodules typically have volume doubling times of 20–400 days, whereas benign nodules usually have volume doubling times of more than 400 days (9–14).

The value of volume doubling time in clinical practice is, however, largely dependent on measurement variability. Two-dimensional volumetric assessments have been found to be unreliable in the detection of volume changes (15,16), while the results obtained with computer-aided three-dimensional (3D) volumetric software have been found to be superior to the results obtained with two-dimensional software (12,17–22).

In a retrospective study, Revel et al (18) reported that software-generated volumetric analyses yielded similar results in nine repeated measurements in 35 (67%) of 52 nodules. Gietema et al (23) found complete agreement for 89% of the 430 solid 15–500-mm³ nodules that were situated entirely within the lung parenchyma. However, not all nodules are purely intraparenchymal. They may also be pleural based, juxtavascular, or fissure attached, and these different locations may influence the accuracy and variability of volumetric measurements. Furthermore, nodule morphology may influence volume measurement variability (23,24).

To our knowledge, quantitative assessment of the independent effects of nodule morphology, size, and location on volume measurement variability has not been performed. Thus, the purpose of our study was to retrospectively assess volume measurement variability in solid pulmonary nodules (volume, 15–500 mm³) detected at lung cancer screening and to quantify the independent effects of nodule morphology, size, and location.

**Materials and Methods**

**Study Group**

Between April 2004 and December 2005, 674 participants underwent baseline low-dose multidetector CT screening for lung cancer. Participants (5917 men, 857 women) were current or former smokers aged 50–75 years with a history of smoking more than 15 cigarettes per day for more than 25 years or more than 10 cigarettes per day for more than 30 years. The prospective screening study was approved by the Dutch Ministry of Health and the medical ethical board at each of the four participating hospitals, and written informed consent was obtained from all participants. Our current retrospective study was performed by using CT data acquired in the prospective screening study. The original approval and informed consent for the prospective study included permission to use data for future retrospective research.

**Imaging**

All examinations were performed by using 16-detector helical CT scanners (Sensation 16, Siemens Medical Systems, Forchheim, Germany; Multischneider IDT or Brilliance 16P, Philips Medical Systems, Cleveland, Ohio), a low radiation dose, and the following parameters: 0.5-second tube rotation, 0.75-mm single-section collimation, 15- or 18-mm table feed per rotation, and 1.3–1.5 pitch. Scanning was performed in a caudocranial direction, without the use of contrast material. Images were obtained from the level of the lung bases (posterior recesses) to the level of the apex with the help of a scout view. Exposure settings (20–30 mAs, 100–140 kVp) depended on the weight of the participant. These settings corresponded to an effective radiation dose of less than 1.6 mSv. We reconstructed axial images with a 1.0-mm thickness at 0.7-mm increments.

**Advances in Knowledge**

- Complete agreement in software-generated volume measurement is dependent on nodule morphology and location and varies from 91% for purely intraparenchymal nodules to 70% for juxtavascular nodules and from 90% for smooth nodules to 34% for irregular nodules.
- Compared with smooth, small, and purely intraparenchymal nodules, irregular and juxtavascular nodules have the largest volume variability; in spiculated, pleural-based, and fissure-attached nodules, variability is moderate; and in lobar and larger (>50 mm³) nodules, variability is small.

**Implication for Patient Care**

- Visual verification of growth, in addition to volumetric measurement, is recommended for pleural-based, fissure-attached, or juxtavascular nodules with ill-defined margins to avoid false-positive results.

**Abbreviations:**

- CI = confidence interval
- OR = odds ratio
- RVD = relative volume difference
- 3D = three-dimensional

**Author contributions:**


Authors stated no financial relationship to disclose.
Nodules were classified as smooth, polylobulated, spiculated, or irregular if it was not smooth, polylobulated, or spiculated. Nodules were further classified as small (15–49 mm³; effective diameter, 3.1–4.6 mm) or large (50–500 mm³; effective diameter, 4.7–9.8 mm), corresponding to two nodule categories (insignificantly small and indeterminate) in our screening program (25). In our retrospective study, we retrieved all information on solid nodules, with the exception of complete and central calcified nodules between 15 and 500 mm³ (effective diameter, 3.1–9.8 mm) detected by both readers in our central database. Nodules larger than 500 mm³ were excluded because the software we used was not developed for use with these larger nodules. We evaluated all CT images. Because two of the 13 local readers manually modified the volume, all of their readings were excluded from our retrospective study; thus, there were 11 local readers. In cases of disagreement between the first and second readers on nodule characteristics, two readers (D.M.Y., Y.W.) retrospectively reevaluated the images to reach consensus.

Statistical Analysis

Because the true nodule volume was unknown, we used the Bland-Altman method (29) to estimate volume measurement variability. The difference between the first and second volume measurements was computed for all nodules and was defined as the absolute volume difference. The relative volume difference (RVD) was defined as the absolute volume difference divided by the mean volume of the two measurements. Volume measurement variability was assessed for both absolute volume difference and RVD, and means and 95% confidence intervals (CIs) were calculated. Thereafter, the absolute value of RVD was calculated because the percentage volume change is directly related to growth (30). Disagreement was defined as small when the absolute value of RVD was less than 5%, moderate when it was between 5% and 15%, and large when it was more than 15%. The percentage of nodules with volume disagreement of more than 25% was com-
computed because 25% was used as the growth criterion in our lung cancer screening program.

To evaluate the influence of nodule characteristics on measurement variability, we performed univariate logistic regression analysis by using volume disagreement as a dependent variable and by using location, morphology, and size as independent variables. Volume disagreement was defined as an absolute value of RVD that did not equal zero, whereas volume agreement was defined as an absolute value of RVD that equaled zero, and odds ratios (ORs) and 95% CIs were computed. In the next step, the independent variables that contributed significantly in the univariate analyses were included in the multivariate logistic regression analysis. Again, ORs with 95% CIs were calculated. Because the definition of volume agreement was rather strict, the following factors were also considered: (a) whether a nodule had a disagreement of more than 5% (absolute value of RVD > 5% or absolute value of RVD ≤ 5%) and (b) whether a nodule had a disagreement of more than 15% (absolute value of RVD > 15% or absolute value of RVD ≤ 15%). All analyses were performed with statistical software (SPSS, version 14.0.1; SPSS, Chicago, Ill.).

**Results**

Of the 6774 participants who underwent baseline screening, 2367 had 4477 solid noncalcified nodules with a volume of 15–500 mm$^3$ detected by both readers. Of these nodules, 252 were excluded because they were evaluated by readers who manually changed the software-generated volume. Thus, we included 4225 nodules in 2239 participants in our analyses. The median volume was 40 mm$^3$ (range, 15–485 mm$^3$). The majority of nodules were purely intraparenchymal (68%, $n = 2853$), had a smooth outer contour (80%, $n = 3375$), and were small (62%, $n = 2603$) (Table 1).

The mean absolute volume difference for all nodules was 0.6 mm$^3$ ± 9.1 (standard deviation) (95% CI: 3.8, 8.0), and the mean RVD was 0.5% (95% CI: −13.4, 14.5%) (Fig 1a). Complete volume agreement between the two readers was recorded in 3646 (86%) nodules. The majority of nodules were purely intraparenchymal (68%, $n = 2853$), had a smooth outer contour (80%, $n = 3375$), and were small (62%, $n = 2603$) (Table 1).

The mean absolute volume difference for all nodules was 0.6 mm$^3$ ± 9.1 (standard deviation) (95% CI: 3.8, 8.0), and the mean RVD was 0.5% (95% CI: −13.4, 14.5%) (Fig 1a). Complete volume agreement between the two readers was recorded in 3646 (86%) nodules. The majority of nodules were purely intraparenchymal (68%, $n = 2853$), had a smooth outer contour (80%, $n = 3375$), and were small (62%, $n = 2603$) (Table 1).

The mean absolute volume difference for all nodules was 0.6 mm$^3$ ± 9.1 (standard deviation) (95% CI: 3.8, 8.0), and the mean RVD was 0.5% (95% CI: −13.4, 14.5%) (Fig 1a). Complete volume agreement between the two readers was recorded in 3646 (86%) nodules. The majority of nodules were purely intraparenchymal (68%, $n = 2853$), had a smooth outer contour (80%, $n = 3375$), and were small (62%, $n = 2603$) (Table 1).

**Table 1**

**RVD and Absolute Volume Difference according to Location, Morphology, and Size in 4225 Solid Pulmonary Nodules Detected at Baseline Screening for Lung Cancer**

<table>
<thead>
<tr>
<th>Nodule Characteristic</th>
<th>No. of Nodules*</th>
<th>Median Volume (mm$^3$)</th>
<th>RVD (%)†</th>
<th>Absolute Volume Difference (mm$^3$)‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Purely intraparenchymal</td>
<td>2853 (68)</td>
<td>36</td>
<td>0.4 ± 5.4</td>
<td>0.5 ± 6.2</td>
</tr>
<tr>
<td>Pleural based</td>
<td>717 (17)</td>
<td>50</td>
<td>1.3 ± 9.4</td>
<td>1.4 ± 15.2</td>
</tr>
<tr>
<td>Juxtavascular</td>
<td>199 (5)</td>
<td>68</td>
<td>0 ± 12.5</td>
<td>0 ± 14.2</td>
</tr>
<tr>
<td>Fissure attached</td>
<td>456 (11)</td>
<td>46</td>
<td>0.2 ± 8.9</td>
<td>0.2 ± 7.9</td>
</tr>
<tr>
<td>Morphology</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smooth</td>
<td>3375 (80)</td>
<td>36</td>
<td>0.4 ± 5.1</td>
<td>0.2 ± 4.6</td>
</tr>
<tr>
<td>Lobulated</td>
<td>611 (14)</td>
<td>64</td>
<td>1.4 ± 1.1</td>
<td>2.1 ± 17.4</td>
</tr>
<tr>
<td>Spiculated</td>
<td>68 (2)</td>
<td>181</td>
<td>0 ± 6.0</td>
<td>−0.7 ± 7.8</td>
</tr>
<tr>
<td>Irregular</td>
<td>171 (4)</td>
<td>40</td>
<td>2.0 ± 16.8</td>
<td>3.6 ± 21.9</td>
</tr>
<tr>
<td>Size</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td>2603 (62)</td>
<td>28</td>
<td>0.2 ± 5.1</td>
<td>0.7 ± 1.6</td>
</tr>
<tr>
<td>Large</td>
<td>1622 (38)</td>
<td>78</td>
<td>0.4 ± 9.5</td>
<td>1.4 ± 14.4</td>
</tr>
</tbody>
</table>

* Data in parentheses are percentages.
† Data are means ± standard deviations.
‡ Data are means ± standard deviations.

In multivariate regression analysis, irregular nodules (Fig 2) had ORs of 15.7 (95% CI: 11.0, 22.2), 12.3 (95% CI: 8.8, 17.7), and 9.6 (95% CI: 6.1, 15.1) for volume discrepancies of more than 0%, more than 5%, and more than 15%, respectively, compared with the smooth reference category. Juxtavascular nodules (Fig 3) had ORs for volume disagreement of 3.5 (95% CI: 2.4, 5.0), 5.5 (95% CI: 3.8, 8.0), and 4.6 (95% CI: 2.8, 7.6), respectively, compared with the purely intraparenchymal reference category. Pleural-based, fissure-attached, spiculated, lobulated, and larger nodules had lower ORs compared with the corresponding reference category of nodules (Table 3).

**Discussion**

Our results show that in 86% of the solid nodules detected in a lung cancer screening trial, complete volume agreement between two readers can be achieved by using semiautomated volumetric software. Large volume discrepancies (absolute value of RVD ≥ 25%) occurred in only 2% of nodules (95% CI: 1.4%, 2.3%). An increase in volume of at least 25% within a 3-month interval was defined as growth in one screening trial (25), and it might have led to a false-positive test result at 3-month follow-up in 2% of the pulmonary nodules...
We also found that irregular and juxtavascular nodules had the largest risk of volume variability; spiculated, fissure-attached, and pleural-based nodules had a moderate risk of volume variability; and lobulated and larger nodules had a small risk of volume variability. Semiautomated software-generated segmentation should be visually verified and manually adjusted if the observer believes segmentation has not been successful, especially with irregular and juxtavascular nodules.

Compared with previous studies on volume measurement variability,
our study not only was larger, but it also covered the complete spectrum of solid nodules detected at baseline screening. To our knowledge, ours was also the only study in which the independent effects of nodule characteristics on volume measurement variability were quantified. Gietema et al (23) used a smaller data set with only purely intraparenchymal nodules. They found complete agreement between two readers regarding a purely intraparenchymal location for 89% of the nodules, which is almost identical to our finding of complete agreement for 91% of the nodules. In the study conducted by Wormans et al (22), two consecutive CT acquisitions were performed within a 10-minute interval in 151 nodules (mean diameter, 7.4 mm) in patients with known pulmonary metastases. The interobserver 95% CI was −5.5% to 6.6%, which was less than that in our study. However, no details on the morphology and location of these nodules were provided. As pulmonary metastases tend to be purely intraparenchymal, smooth, and round, the conclusion reached by Wormans et al (22) might be applicable to only that type of nodule. In the other studies (13,18,21) on volume measurement error published thus far, the number of nodules ranged from 13 to 62. The nodules were usually smaller than 10 mm and incidentally found on routine CT images. Because different volumetric software programs were used, comparison with our results is impossible.

An explanation for the observed volume disagreement between the two readers is related to the fact that semiautomatic volumetric measurements may vary according to the positioning of the seed point by the observer, which is the only nonautomated part of the procedure. When a spherical 3D template gradually expands from the seed point, different starting positions within the nodules may lead to different volumetric results. In nonsmooth nodules, it is difficult to identify the correct borderline, and in nonspherical nodules, the segmentation tends to vary with the seed point owing to the spherical 3D template used in the software program. In pleural-based, fissure-attached, and juxtavascular nodules, both measurement errors occur, whereas in large nodules, the larger disposable seed point area leads to more measurement variability. The software program used in our study provides a postprocessing option for unsatisfactory segmentation. Two (Y.W., D.M.X.) of the original 13 readers in our study were accustomed to changing the segmentation in cases of an unsatisfactory software-generated volume. The nodules these two readers evaluated were excluded because they performed the first reading and their results were not comparable to the results of the second readers, who used only the semia-
tomatic measurement and could have introduced additional variability. The number of nodules evaluated by these two readers ($n = 252$) was relatively small compared with the total number of nodules ($n = 4477$).

A limitation of our study was that there was no reference standard for the true nodule volume. However, in practice, this is impossible to achieve with noninvasive means. Furthermore, we wanted to assess volume measurement variability between two independent readers. Another limitation of our study was that our results are applicable only for the specific software used, although they may be applicable for other software if identical procedures are used to measure the volume. Furthermore, we evaluated only the variability between two observers and not the variability between two successive CT examinations, which may be considerably larger and relevant to the assessment of volumetric growth over time.

In conclusion, semiautomated software-generated volumetric measurement was completely reproducible in 86% of the solid 15–500-mm$^3$ nodules. Nodule morphology, location, and size influenced volume measurement variability, particularly among the juxtavascular and irregular nodules, in which visual verification of growth is recommended to avoid false-positive results.

Acknowledgments: We thank our colleagues from the Department of Radiology, University Medical Center Groningen, University of Groningen, especially Yingru Zhao, MD, and Fen- neke van de Wiel, MD, for their support in the evaluation of the nodules; Dr Estelle J. K. Noach, MD, PhD, for her critical review of the manuscript; and Dr Peter Paul Taverner, MD, Dr Peter van Oogen, PhD, and Wim Tkuker, PhD, for their advice and technical support.

References