

Volume versus diameter assessment of small pulmonary nodules in CT lung cancer screening

Daiwei Han, Marjolein A. Heuvelmans, Matthijs Oudkerk

University of Groningen, University Medical Center Groningen, Center for Medical Imaging–North East Netherlands, Groningen, the Netherlands

Contributions: (I) Conception and design: All authors; (II) Administrative support: None; (III) Provision of study materials or patients: None; (IV) Collection and assembly of data: All authors; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Marjolein A. Heuvelmans, MD, PhD. CMI-Center for Medical Imaging EB45, University Medical Center Groningen, Hanzeplein 1, 9713 GZ Groningen, the Netherlands. Email: m.a.heuvelmans@umcg.nl.

Abstract: Currently, lung cancer screening by low-dose chest CT is implemented in the United States for high-risk persons. A disadvantage of lung cancer screening is the large number of small-to-intermediate sized lung nodules, detected in around 50% of all participants, the large majority being benign. Accurate estimation of nodule size and growth is essential in the classification of lung nodules. Currently, manual diameter measurements are the standard for lung cancer screening programs and routine clinical care. However, European screening studies using semi-automated volume measurements have shown higher accuracy and reproducibility compared to diameter measurements. In addition to this, with the optimization of CT scan techniques and reconstruction parameters, as well as advances in segmentation software, the accuracy of nodule volume measurement can be improved even further. The positive results of previous studies on volume and diameter measurements of lung nodules suggest that manual measurements of nodule diameter may be replaced by semi-automated volume measurements in the (near) future.

Keywords: Semi-automated volume measurement; manual diameter measurement; small pulmonary nodule

Submitted Oct 07, 2016. Accepted for publication Dec 21, 2016.

doi: 10.21037/tlcr.2017.01.05

View this article at: <http://dx.doi.org/10.21037/tlcr.2017.01.05>

Introduction

Lung cancer is the leading cause of cancer-related mortality in men and women worldwide (1). This is mainly due to the fact that lung cancer patients are mostly asymptomatic in early stages. Patients presenting with symptoms such as cough or chest pain in the clinics often already have advanced lung cancer with very limited survival time despite of treatment. Therefore, early detection of lung cancer is extremely important.

The National Lung Screening Trial (NLST), a large-scale randomized controlled trial including over 53,000 participants, has demonstrated that low-dose computed tomography (LDCT) screening for individuals at high risk for lung cancer, reduced lung cancer mortality by 15–20%, when compared to chest radiographs (2). The result of NLST was translated by several U.S. medical associations,

including the U.S. Preventive Service Task Force, into recommendation for lung cancer screening using LDCT for high-risk individuals (3,4). Since February 2015, lung cancer screening for high risk individuals is covered by Medicare (5).

Despite of the benefits regarding lung cancer related mortality reduction brought by LDCT lung cancer screening, a major drawback of lung cancer screening is its high rate of false-positive screen results. A challenging problem in lung cancer screening is the high prevalence of small-to-intermediate sized (<500 mm³ or <10 mm) lung nodules found in LDCT lung cancer screening. Up to 66% of participants enrolled for LDCT lung cancer screening have at least one lung nodule, with the large majority being benign (6). Across all three screening rounds of the NLST, 24.2% of participants received a positive screen

result, 96.4% of these were false-positive (2). Although most nodules were classified as benign by non-invasive follow up scans, false-positive screening results may cause anxiety and unnecessary invasive diagnostic procedures in some participants, which may come with complications and increased healthcare cost. In order to reduce the false-positive rate, a method to accurately identify malignant nodules is required.

Diameter measurements are widely used in CT lung cancer screening. Due to its simplicity, it is easily applicable and it is currently the main method worldwide for measuring lung nodule size. A number of medical associations have incorporated diameter measurement in their recommended nodule management protocol, including the Fleischner society lung nodule recommendations and the Lung CT screening reporting and data system (Lung-RADS) (7,8). With the advent of thin slice CT, more than a decade ago, as well as the availability of three-dimensional segmentation software, semi-automated volume measurements of lung nodules have become a valuable alternative option. The Dutch-Belgian lung cancer screening trial (NELSON) is the first large scale lung cancer screening trial that based its nodule management protocol on semi-automatically measured volume of lung nodules instead of manually measured diameter, and on nodule growth in terms of volume-doubling time. Implementation of an extra screen result for indeterminate lung nodules (volume 50–500 mm³ or volume-doubling time (VDT) 400–600 days at incidence screens) leading to an extra short term LDCT, had led to a far lower false-positive rate compared to NLST (1.7% *vs.* 26.6%), a comparable sensitivity (92.5% *vs.* 93.5%) and a significantly higher specificity (98.3% *vs.* 73.4%) at baseline (2,9).

Several nodule characteristics were shown to be associated with lung cancer. Lung nodule margin may also help with differentiation of malignant from benign nodules. Spiculated and lobulated nodules have higher probability of being malignant compared to round nodules with smooth margin (10-13). However, nodule size and nodule growth rate were found to be most important predictors (14-19).

In lung cancer screening management protocols, nodules are divided to different risk groups by diameter or volume-based cutoffs for new nodules, and nodule growth or growth rate cutoffs for existing ones. Accurate nodule size estimation is necessary for correct classification of nodules to their risk groups. Misclassification of a nodule could potentially result in under- or overestimation of its malignancy probability, misdiagnosis or over-diagnosis of

lung cancer, along with increased frequency of unnecessary invasive diagnostic procedures.

This review first focuses on the two largest CT lung cancer screening studies that have used diameter or volume measurements for the estimation of lung nodule size. Secondly, inter- and intrareader variability studies for both volume and diameter measurement will be discussed. Thirdly, an overview of recent phantom studies focusing on the accuracy and precision for volume- and diameter based measurement will be provided. Fourthly, the influence of CT-scan and reconstruction parameters on volume measurements will be discussed. Finally, future perspectives on lung nodule measurements will be proposed.

Lung cancer screening

In the NLST, nodule management was based on manually-measured maximum nodule diameter, or diameter growth on follow-up CTs. Participants with nodules of ≥ 4 mm or with increase in diameter of at least 10% were considered screen positive (20). However, for a nodule 4 mm in diameter, a 10% increase would mean 0.4 mm, which lies within the range of the 95% confidence interval of manual measurements of small nodules.

To reduce the high false positive rate of the NLST, the American College of Radiology released Lung CT Screening Reporting and Data System (Lung-RADS), a classification system for LDCT lung cancer screening. According to Lung-RADS, lung nodules are classified based on maximum diameter for round nodules and based on mean diameter measurement (mean of maximum transverse diameter and perpendicular diameter) for non-round nodules, instead of the use of only maximum diameter in NLST. The diameter cut-off for benign nodules has been increased from 4 to 6 mm. Two intermediate categories, which are also considered as positive result were added, a 6–8 mm category (6-month follow-up) and an 8–15 mm category (3-month follow-up). Growth has been defined as a 1.5 mm increase in diameter, instead of the 10% criterion in the NLST protocol. A growing nodule will be moved up from its original category to a higher risk category, requiring more intensive management. Additionally, nodules presenting with additional imaging findings such as spiculation, ground glass nodule that doubles in size within a year, as well as enlarged lymph nodes may be classified into a higher risk group.

The increase of cut-off for benign nodules may decrease the sensitivity for malignant nodules. In a retrospective

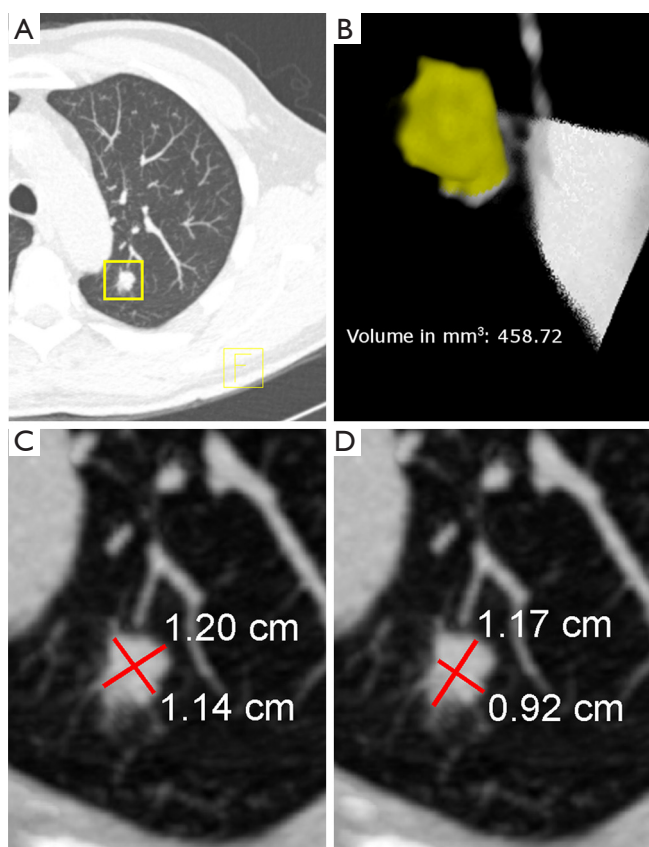


Figure 1 Transverse images of a solid pulmonary nodule and its location (A), semi-automatically assessed volume (B) and two possible manual diameter measurements (C) and (D).

study, where Pinsky *et al.* applied Lung-RADS criteria to the NLST protocol, it was found that Lung-RADS protocol showed a higher specificity when compared to the NLST protocol (87.2% *vs.* 73.4%). However, the lung cancer sensitivity of Lung-RADS is lower than NLST (84.9% *vs.* 93.5%) (3). In another study, where Lung-RADS was applied to a total of 2,180 screening subjects with high-risk for lung cancer, of which 577 (26%) of patient's clinical follow-up was unavailable, McKee *et al.* have shown that reclassifying nodules between 4 mm and 6 mm in diameter from a positive to a negative screen result did not cause an increase in false-negative results (4). However, in this study the number of nodules that were reclassified, only 152, was limited. Given in NLST, nodules of this size category (4–6 mm) had a positive predictive value for lung cancer of 0.5% (2), and 0.8 out of the 152 nodules predicted to be false positive. Therefore, more research is needed to confirm these results.

The NELSON trial is a large-scale randomized-

controlled lung cancer screening trial investigating whether LDCT can reduce lung cancer related mortality by 25% compared to no screening at 10-year follow-up. One of the main differences between NELSON and NLST is the fact that whereas NLST used nodule diameter as its indicator to assess nodule size and growth, NELSON used semi-automatically measured nodule volume. In contrast to NLST, which followed a black and white approach in classifying nodules, NELSON introduced an intermediate screen result (volume 50–500 mm³) for nodules with highest uncertainty in their nature, in addition to a negative screen result (volume <50 mm³) and a positive screen result (volume >500 mm³). Screenees with an indeterminate result received a short-term follow-up by low-dose chest CT. Nodules that had grown by 25% in volume at the short term follow-up CT, and had a VDT <400 days, were considered to be positive. Although the final results of the NELSON trial are still awaited, it was found that the NELSON strategy led to a much higher positive predicted value compared to diameter-based protocols, with comparable lung cancer sensitivity and negative predictive value (9).

Inter- and intrareader variability of nodule volume and diameter

Inter- and intrareader variability studies have played an important role in the development of reliable diagnostic tools, and understanding of the variability of screening outcomes in the field of lung cancer screening. A measurement method that yields low inter- and intravariability results in consistent treatment recommendations. For this, a measurement needs to be objective and has to be an accurate representation of the true size of the object that is being measured. It is therefore logical to assume that estimating nodule size by semi-automated volume measurements is a superior method compared to manual diameter measurements (*Figure 1*). Since there is limited number of studies that compare inter-and intrareader variability of small-sized lung nodule measurements, studies comparing volume- and diameter measurements of larger lesions will also be discussed.

In a study where 54 solid nodules were measured diameter wise by multiple radiologists, Revel *et al.* found that for intrareader variability, 1.6 mm growth cut-off would ensure the detection of true growth, while for multiple radiologists a 1.7 mm growth cut-off would ensure the detection of true growth (21). This study was performed with a single CT scan. It is known that lung nodules are

rarely symmetrical or spherical in shape, and the variation of inspiration level during CT scanning can cause nodules to be scanned at a differing rotational position, thus adding additional variability to diameter measurements. “Coffee break” studies, where measurement variability is evaluated after repeat CT scans within 15 minutes, can demonstrate the inherent variability of inspirational variation. In a study where 30 non-small cell lung cancers with mean lesion size of 3.7 cm (range: 1.0–8.0 cm) were evaluated for measurement variability using repeat CT scans within 15 minutes, the 95% limits of agreement was found to be ± 4.8 mm, and 33% of nodules had measurement differences between radiologists greater than 2 mm (22). Yet, several other factors may have contributed to the large variability found in this study. Firstly, tumors in this study were much larger in size compared to the study by Revel *et al.* Although not statistically significant ($P=0.06$), Oxnard *et al.* have found that larger lesions tended to have larger measurement variability, when compared to smaller nodules. Secondly, the morphology of these 30 tumors had not been described, as non-smooth lesions may yield larger measurement variation compared to smooth nodules. Nevertheless, the result of these studies questions the reproducibility of manual diameter measurements of lung nodules.

Zhao *et al.* evaluated the inter- and intra-scan variability of 32 lesions [mean size: 37.2 mm (range, 10.7–81.5 mm)] of three radiologists (23). The radiologists were blinded and the order of the scans were randomized. For the evaluation of measurement repeatability, the first and the second radiologist had 2 days between two repeat reading sessions, while for the third radiologist all readings were done in one session. The 95% limits of agreement for unidimensional intra-scan and inter-scan variability varied greatly among the three radiologists and were $\pm 10.2\%$ and $\pm 16.9\%$ for the first radiologist, $\pm 19.2\%$ and $\pm 21.8\%$ for the second radiologist, and $\pm 25.9\%$ and $\pm 22.9\%$ for the third radiologist. Interestingly, the third radiologist who did all the readings in one session, did not show better repeatability compared to other radiologist. Although all of the 95% limits of agreement from this study fell within the guideline of Response Evaluation Criteria in Solid Tumors (RECIST) where 30% diameter increase is considered as significant growth, this study showed that measurement variability of manual diameter measurements varies greatly between readers.

The variability of diameter measurements is influenced by repeat scans. Due to the difference in inspirational level, the maximum transverse diameter of an asymmetrical lung nodule

can vary greatly between two consecutive scans because of the difference in nodule rotation. Therefore, it is reasonable to assume the volume of the nodule to stay constant, while maximum diameter might be different on repeat scans. In a study where Gietema *et al.* evaluated interobserver variability of semi-automated volume measurements (24), 47/430 (11%) nodules had interobserver variability while the rest of the lung nodules 383/430 (89%) did not. The 95% limits of agreement of these 47 nodules in which interobserver variability occurred, was found to be $\pm 25.2\%$. In another study, Gietema *et al.* evaluated the influence of lung nodule size, inspiration level, and nodule morphology on volume measurement variability of repeat scan within 15 minutes while using the same volume measurement software (25). The 95% limits of agreement for a total of 218 nodules was $\pm 22.5\%$, which is comparable to $\pm 25.2\%$ found for nodules that had interobserver variability in the previous study. For nodules that were completely segmented by measurement software, the inspiration level was only weakly related to the measurement precision. The influence of nodule shape on the measurement precision was found to be significant ($P<0.001$), considering spherical nodules and nonspherical nodules had 95% limits of agreement of $\pm 12\%$ and $\pm 28.4\%$, respectively. It was suggested to set the growth cut-off for semi-automated volume measurements of spherical nodules at 15% whereas for nonspherical nodules to be set at 30%. However, nonspherical nodules can also be subdivided to lobulated, spiculated, and irregular based on nodule margin, each may influence on measurement variability by different extent. By adopting separate growth cut-off for each nodule morphological categories more optimal nodule management in lung cancer screening might be achieved.

Several other “coffee break” studies have found similar limits of agreement for automated nodule volume measurements. In the study by Wormanns *et al.* the 95% limit of agreement was $\pm 21.3\%$ (26), while Goodman *et al.* reported a 95% limit of agreement of $\pm 25.6\%$ (27). Hein *et al.* compared the interscan interreader variability between standard-dose CT and ultra-low dose CT, and reported maximum 95% limits of agreement of $\pm 27\%$ (28). The similarity of this value compared to values reported in the literature for low-dose CT scanners suggests that semi-automated volume measurement at ultra-low dose is feasible without compromising the precision of volume measurement.

One important point is that volume measurement works conjointly with its semi-automatic function. The benefit of volume measurement alone cannot be clearly seen through

the comparison between manual diameter measurement and semi-automated measurement, as it is not known how much does the automatic function plays part. By comparing semi-automated diameter measurement with semi-automated volume measurement, the true benefit of volume measurement can be shown. Zhao *et al.* have compared the semi-automated diameter measurement with semi-automated volume measurement of lung tumors, where he concluded that growth cut-off, based on the 95% limit of agreement, for semi-automated diameter measurement should be set at 8%, while for semi-automated volume measurement it should be set at 15%. Since the mean tumor size in this study varied from 34 to 37 mm in diameter depending on the reader, for a hypothetical spherical tumor of 35 mm in diameter assuming the tumor to be a perfect sphere, a variation of 8% in diameter, would result in measurement variability of around 25% in volume compared to the 15% cutoff for volume measurements as set by the authors. The low measurement variability compared to other studies found in literature may be pointing to the relation between nodule/tumor size and measurement variability. A disadvantage of comparison between two types of semi-automated measurements is that the performance can vary between measurement softwares. Since it is unknown how the two software programs in the above mentioned studies perform compared to other commercially available softwares, one needs to be cautious when making conclusion of which type of measurement has better accuracy and precision.

Studies discussed so far focused on measurement of static lung lesions. Studies of lung lesions that are tested to be sensitive to targeted therapy can help to demonstrate the sensitivity and specificity of volume and diameter measurement at detecting size change of lesions. Zhao *et al.* compared the sensitivity of diameter and volume measurements in detecting size shrinkage of lung lesions as a response to gefitinib treatment, for lung tumors with and without EGFR sensitizing mutations (29). The optimal threshold for detecting response to gefitinib treatment were determined to be 7.0% and 24.9% in decrease in diameter and volume, respectively. For diameter measurements, the sensitivity for detecting the response of EGFR sensitizing mutations was 71% while specificity was 78%, whereas for volume measurements, the sensitivity was 90% and specificity 89%. Although the study focused on the measurement of lung masses instead of lung nodules, the superiority in sensitivity and specificity of volume measurement in detecting change in size of tumors suggests

similar performance difference also applies for smaller lung nodules.

In summary, although limit of agreement for both manual diameter measurements and semi-automated volume measurements for the studies discussed above lies in the same range in terms of absolute percentages, the percentage of lung nodules in which an actual inter-reader difference found was with 11% far lower for semi-automated nodule volume measurements compared to manual diameter measurements, where inter-reader variability occurs commonly. Furthermore, the extra dimension in volume measurements should be taken into account. Assuming a nodule as a sphere, a variation of 20% in nodule diameter, which is commonly found in the different studies, refers to a variation of 72% in nodule volume.

Phantom studies

Up until now, static and dynamic lung nodules and tumors have been discussed. However, a disadvantage of patient studies is that the true size of lung nodules is not known. CT phantom studies, although less representative of CT scans in the clinical setting, allows CT-derived nodule size assessment of artificial nodules to be compared with their true size. In addition, phantom studies allow optimization of CT scan and reconstruction parameters without harming test subjects with radiation exposure.

Semi-automated volume measurements were compared to manual diameter measurements in an anthropomorphic phantom study by Xie *et al.* (30). It was reported that low-dose CT yielded more accurate volume measurement when using a semi-automated method than using a manual method. In their anthropomorphic study of measuring 15 spherical nodules of 3, 5, 8, 10, and 12 mm in diameter (corresponding to 14, 65, 268, 523 and 904 mm³), both semi-automated volume measurement and volume derived from manual diameter measurement significantly underestimated the size of nodules of three different densities (−800, −630, +100 HU). For solid nodules (+100 HU), semi-automated volume measurements had significantly smaller underestimation when compared to manual diameter derived volume (7.6±8.5% vs. 26.4±15.5%). Furthermore, manual diameter measurements had significantly larger underestimation compared to semi-automated diameter measurements (9.2±6.0% vs. 3.7±7.1%). In addition to this, solid nodules that were smaller had larger relative underestimation in CT-derived volume measurements and diameter measurements,

when compared to larger nodules. However, there are contradicting findings from previous studies, as some have reported a similar trend of underestimation with decreasing nodule size (31,32), while others reported increasing overestimation with decreasing volume size (33-38). The so-called partial volume effect has been considered as an important contributor to potential error in semi-automated volume measurements of lung nodules. Most softwares use voxel counting method to calculate nodule volume. A problem arises when voxels consist of partly nodule tissue and partly lower density lung parenchyma. As a result, voxels around the margin of nodules would present as the average density of nodule tissue and parenchyma in relative radiodensity, seen as blurry margin. Depending on the threshold set by the measurement software, these partial-volume voxels can either be counted as part of nodule volume or can be excluded, hence resulting either over- or underestimation of nodule volume.

In a subsequent phantom study, Xie *et al.* evaluated semi-automated volume measurements of small irregular lung nodules (spiculated and lobulated ones) ranging from 5.1 ± 0.2 to 88.4 ± 3.4 mm³ in volume and -51 ± 11 to $+157 \pm 8$ HU in actual density (39). The CT-derived semi-automated volume of irregular nodules was significantly smaller than the actual volumes with a mean underestimation of 18.9 ± 11.8 mm³ (percentage underestimation: $39\% \pm 21\%$). The percentage underestimation for irregular nodules was found to be larger than for spherical nodules evaluated in the previous study ($39 \pm 21\%$ vs. $7.6 \pm 8.5\%$). Additionally, actual nodule volume and nodule shape were found to significantly influence CT-derived volume. However, it was found that the observer did not influence the CT-derived volume. The fact that nodules that are smaller and with complicated margin have larger measurement error compared to larger and spherical nodules, can be explained by the surface area to volume ratio (SA/V), since nodules with large SA/V ratio are more under influence of partial volume effect than nodules with low SA/V ratio such as large spherical nodules.

In summary, in phantom studies it was found that semi-automated volume measurements led to a more accurate estimation of nodule size as manual diameter measurements. However, both measurement techniques led to an underestimation of actual nodule size.

Effect of scan and reconstruction parameters

Optimization of scan and reconstruction parameters can

improve the accuracy of lung nodule measurements. Several parameters have been examined on their influence on volume measurement accuracy and precision, including slice thickness and reconstruction kernel.

As explained earlier, partial volume effect can cause volume measurement errors especially in nodules small in size and with complex margins, as they have large SA/V. However, by reducing the slice thickness of CT, the voxel size can be decreased. Therefore, the influence of the partial volume effect can be reduced and measurement accuracy can be improved. The influence of slice thickness on volume measurement has been evaluated in a number of studies. Winer-Muram *et al.* have found a difference of 20% in volume measurement of tumors between thin (2–3 mm) and thick (8–10) slice thickness settings (38). Similar findings have also been reported by Petrou *et al.* (40) and Kuhnigk *et al.* (41).

In phantom studies, the influence of slice thickness on semi-automated volume measurement has been evaluated as well. In a recent study, Li *et al.* evaluated 40,000 volume measurements of 48 nodules ranging from 5 to 20 mm in size to determine factors that contribute substantially to measurement errors of lung nodule size (42). Slice thickness \times collimation along with nodule size and attachment to vessels and chest wall were found to be the main factors of measurement error. Similar findings regarding slice thickness also have been reported by Winer-Muram *et al.* in which volume overestimation varied directly with section thickness (38).

In CT-imaging a compromise exists between image detail and image noise. During image reconstruction, mathematical filters, also known as reconstruction kernels, are applied to change some of the image characteristics. Detailed kernels, also known as hard kernels, are selected to enhance detail, usually at the expense of increased imaging noise, while soft kernels are selected to reduce noise in an image. However, reduction of noise usually increases blurring and reduces visibility of detail. Like the blurring caused by partial volume effect, which has an influence on nodule volume measurements, soft kernels can also influence nodule volume measurements in a similar way, by introducing blurring along the nodule margin. Ko *et al.* reported that the choice of hard or soft kernels had a significant effect on the measurement error of synthetic nodules (43). Similarly, Li *et al.* reported hard kernels yielded better repeatability coefficient than those images reconstructed using medium kernels (42).

Although hard kernels have generally yielded better

measurement results, with reduced measurement error and better precision than soft kernels, image noise remains to be a problem and can possibly affect lung nodule detectability and volume measurement. Punwani *et al.* reported decreasing sensitivity of nodule detection (>4 mm) at signal-to-noise ratio (SNR) of 1.5 and lower (44), with similar finding also reported by Li *et al.* (45). Nevertheless, Christie *et al.* reported no significant impact of noise on volume measurement (46). However, Xie *et al.* reported underestimation of nodule measurements at lower doses (30). The latter was confirmed by Willeminck *et al.*, reporting an underestimation of up to 23.9% for nodules >5 mm (47). The difference in reported results could be caused by the choice of different slice thicknesses and type of reconstruction kernel.

Future perspectives

There is an accumulating evidence that shows volume measurements of lung nodules is more accurate and reproducible when compared to diameter measurements. Through the optimization of scanning and reconstruction parameters, the accuracy and reproducibility of measurement can be further improved, reaching even closer to the physical volume of lung nodules. However, this may be lesser the case for diameter measurement.

Recently, iterative reconstruction has been introduced to CT imaging. Compared to reconstruction kernels, iterative reconstruction aims to reduce the noise of images while minimising the compromise to image detail. This allows the possibility of ultra-low dose CT-scanning (1/10 of radiation dose compared to a regular CT scan), producing CT images of image quality and detail comparable to LDCT. Sui *et al.* reported that ultra-low dose CT with iterative reconstruction had higher sensitivity in detecting lung nodules (92.1% and 92.9% for reader 1 and 2, respectively) compared to LDCT with filtered back projection (88.9% and 86.6% for radiologist 1 and 2, respectively). Furthermore, when iterative reconstruction is compared to LDCT with filtered back projection, no significant differences in nodule volume or diameter measurements were found (48).

Commercially available software can vary greatly in measurement accuracy and precision, not to mention the poor performance in measuring the size of sub-solid/part-solid nodules, due to their lower contrast ratio to the lung parenchyma. Zhao *et al.* reported significant differences in volumetric measurements from the comparison of

three commercially available volume measurement softwares which could affect the classification of nodules based on their size (49). In addition to this, one of the main factors for large variability in semi-automatic volume measurement of lung nodules is vessel and pleura attachment. Therefore, the improvement of segmentation software as well as standardization of CT imaging parameters is an important part in the implementation of semi-automated volume measurement into clinical practice.

The final screen results of the NELSON trial are awaited before the decision on implementation of lung cancer screening in Europe can be made. However, with the implementation of lung cancer screening in the United States, and the high number of incidentally detected nodules in clinical care worldwide, it is of major importance to use a precise and accurate manner to estimate nodule size and detect nodule growth. The positive results of previous studies suggest that manual measurements of nodule diameter may be replaced by semi-automated volume measurements in the (near) future.

Summary

We have reviewed the comparison of semi-automated volume measurement and diameter measurement of lung nodules. Although manual diameter measurements are currently the standard in U.S. lung cancer screening programs and are used for lung nodules detected in routine clinical care, results of European screening studies using semi-automated volume measurements in terms of false-positive screen results and positive predictive value may not be neglected. There is an accumulating evidence that semi-automatic volume measurements have higher accuracy and reproducibility compared to diameter measurements. Furthermore, with optimization of CT scanners and reconstruction parameters, and advancement in semi-automated volume measurement software, the accuracy and reproducibility of volume measurements can be improved even further. The positive results of previous studies on volume and diameter measurements of lung nodules suggest that manual measurements of nodule diameter may be replaced by semi-automated volume measurements in the (near) future.

Acknowledgements

None.

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

References

1. Previous Version: SEER Cancer Statistics Review, 1975-2011. Available online: https://seer.cancer.gov/archive/csr/1975_2011/
2. National Lung Screening Trial Research Team, Aberle DR, Adams AM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011;365:395-409.
3. Pinsky PF, Gierada DS, Black W, et al Performance of Lung-RADS in the National Lung Screening Trial: a retrospective assessment. *Ann Intern Med* 2015;162:485-91.
4. McKee BJ, Regis SM, McKee AB, et al. Performance of ACR Lung-RADS in a Clinical CT Lung Screening Program. *J Am Coll Radiol* 2016;13:R25-9.
5. Medicare Coverage for Lung Cancer Screening Frequently Asked Questions. Available online: <http://www.lung.org/lung-health-and-diseases/lung-disease-lookup/lung-cancer/learn-about-lung-cancer/lung-cancer-screening/medicare-coverage-faq.html>
6. Swensen SJ, Jett JR, Sloan JA, et al. Screening for lung cancer with low-dose spiral computed tomography. *Am J Respir Crit Care Med* 2002;165:508-13.
7. 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.
8. Lung CT Screening Reporting and Data System (Lung-RADS™). Available online: <https://www.acr.org/Quality-Safety/Resources/LungRADS>
9. Horeweg N, van Rosmalen J, Heuvelmans MA, et al. Lung cancer probability in patients with CT-detected pulmonary nodules: a prespecified analysis of data from the NELSON trial of low-dose CT screening. *Lancet Oncol* 2014;15:1332-41.
10. Wahidi MM, Govert JA, Goudar RK, et al. Evidence for the treatment of patients with pulmonary nodules: when is it lung cancer?: ACCP evidence-based clinical practice guidelines (2nd edition). *Chest* 2007;132:94S-107S.
11. Aoki T, Nakata H, Watanabe H, et al. Evolution of peripheral lung adenocarcinomas: CT findings correlated with histology and tumor doubling time. *AJR Am J Roentgenol* 2000;174:763-8.
12. Yankelevitz DE, Reeves AP, Kostis WJ, et al. Small pulmonary nodules: volumetrically determined growth rates based on CT evaluation. *Radiology* 2000;217:251-6.
13. Winer-Muram HT, Jennings SG, Tarver RD, et al. Volumetric growth rate of stage I lung cancer prior to treatment: serial CT scanning. *Radiology* 2002;223:798-805.
14. Henschke CI, Yankelevitz DE, Naidich DP, et al. CT screening for lung cancer: suspiciousness of nodules according to size on baseline scans. *Radiology* 2004;231:164-8.
15. Gould MK, Donington J, Lynch WR, et al. Evaluation of individuals with pulmonary nodules: when is it lung cancer? Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest* 2013;143:e93S-120S.
16. Xu DM, van Klaveren RJ, de Bock GH, et al. Role of baseline nodule density and changes in density and nodule features in the discrimination between benign and malignant solid indeterminate pulmonary nodules. *Eur J Radiol* 2009;70:492-8.
17. Gould MK, Ananth L, Barnett PG, et al. A clinical model to estimate the pretest probability of lung cancer in patients with solitary pulmonary nodules. *Chest* 2007;131:383-8.
18. McWilliams A, Tammemagi MC, Mayo JR, et al. Probability of cancer in pulmonary nodules detected on first screening CT. *N Engl J Med* 2013;369:910-9.
19. Xu DM, van Klaveren RJ, de Bock GH, et al. Limited value of shape, margin and CT density in the discrimination between benign and malignant screen detected solid pulmonary nodules of the NELSON trial. *Eur J Radiol* 2008;68:347-52.
20. Dingemans A-MC, Reck M, Westeel V. Lung Cancer. European Respiratory Society. Wakefield: Charlesworth Press 2015:293.
21. Revel MP, Bissery A, Bienvenu M, et al. Are two-dimensional CT measurements of small noncalcified pulmonary nodules reliable? *Radiology* 2004;231:453-8.
22. Oxnard GR, Zhao B, Sima CS, et al. Variability of lung tumor measurements on repeat computed tomography scans taken within 15 minutes. *J Clin Oncol* 2011;29:3114-9.
23. Zhao B, James LP, Moskowitz CS, et al. Evaluating variability in tumor measurements from same-day repeat CT scans of patients with non-small cell lung cancer. *Radiology* 2009;252:263-72.

24. Gietema HA, Wang Y, Xu D, et al. Pulmonary nodules detected at lung cancer screening: interobserver variability of semiautomated volume measurements. *Radiology* 2006;241:251-7.
25. Gietema HA, Schaefer-Prokop CM, Mali WP, et al. Pulmonary nodules: Interscan variability of semiautomated volume measurements with multisection CT-- influence of inspiration level, nodule size, and segmentation performance. *Radiology* 2007;245:888-94.
26. 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.
27. 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;186:989-94.
28. Hein PA, Romano VC, Rogalla P, et al. Variability of semiautomated lung nodule volumetry on ultralow-dose CT: comparison with nodule volumetry on standard-dose CT. *J Digit Imaging* 2010;23:8-17.
29. Zhao B, Oxnard GR, Moskowitz CS, et al. A pilot study of volume measurement as a method of tumor response evaluation to aid biomarker development. *Clin Cancer Res* 2010;16:4647-53.
30. Xie X, Zhao Y, Snijder RA, et al. Sensitivity and accuracy of volumetry of pulmonary nodules on low-dose 16- and 64-row multi-detector CT: an anthropomorphic phantom study. *Eur Radiol* 2013;23:139-47.
31. Linning E, Daqing M. Volumetric measurement pulmonary ground-glass opacity nodules with multi-detector CT: effect of various tube current on measurement accuracy--a chest CT phantom study. *Acad Radiol* 2009;16:934-9.
32. Zhang L, Yankelevitz DF, Henschke CI, et al. Zone of transition: a potential source of error in tumor volume estimation. *Radiology* 2010;256:633-9.
33. Das M, Ley-Zaporozhan J, Gietema HA, et al. Accuracy of automated volumetry of pulmonary nodules across different multislice CT scanners. *Eur Radiol* 2007;17:1979-84.
34. Kim KG, Goo JM, Kim JH, et al. Computer-aided diagnosis of localized ground-glass opacity in the lung at CT: initial experience. *Radiology* 2005;237:657-61.
35. Marten K, Funke M, Engelke C. Flat panel detector-based volumetric CT: prototype evaluation with volumetry of small artificial nodules in a pulmonary phantom. *J Thorac Imaging* 2004;19:156-63.
36. Tao P, Griess F, Lvov Y, et al. Characterization of small nodules by automatic segmentation of X-ray computed tomography images. *J Comput Assist Tomogr* 2004;28:372-7.
37. Way TW, Chan HP, Goodsitt MM, et al. Effect of CT scanning parameters on volumetric measurements of pulmonary nodules by 3D active contour segmentation: a phantom study. *Phys Med Biol* 2008;53:1295-312.
38. Winer-Muram HT, Jennings SG, Meyer CA, et al. Effect of varying CT section width on volumetric measurement of lung tumors and application of compensatory equations. *Radiology* 2003;229:184-94.
39. Xie X, Willeminck MJ, de Jong PA, et al. Small irregular pulmonary nodules in low-dose CT: observer detection sensitivity and volumetry accuracy. *AJR Am J Roentgenol* 2014;202:W202-9.
40. Petrou M, Quint LE, Nan B, et al. Pulmonary nodule volumetric measurement variability as a function of CT slice thickness and nodule morphology. *AJR Am J Roentgenol* 2007;188:306-12.
41. 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;25:417-34.
42. Li Q, Gavrielides MA, Sahiner B, et al. Statistical analysis of lung nodule volume measurements with CT in a large-scale phantom study. *Med Phys* 2015;42:3932-47.
43. Ko JP, Rusinek H, Jacobs EL, et al. Small pulmonary nodules: volume measurement at chest CT--phantom study. *Radiology* 2003;228:864-70.
44. Punwani S, Zhang J, Davies W, et al. Paediatric CT: the effects of increasing image noise on pulmonary nodule detection. *Pediatr Radiol* 2008;38:192-201.
45. Li X, Samei E, Barnhart HX, et al. Lung nodule detection in pediatric chest CT: quantitative relationship between image quality and radiologist performance. *Med Phys* 2011;38:2609-18.
46. Christe A, Szucs-Farkas Z, Huber A, et al. Optimal dose levels in screening chest CT for unimpaired detection and volumetry of lung nodules, with and without computer assisted detection at minimal patient radiation. *PLoS One* 2013;8:e82919.
47. Willeminck MJ, Leiner T, Budde RP, et al. Systematic error in lung nodule volumetry: effect of iterative reconstruction versus filtered back projection at different CT parameters. *AJR Am J Roentgenol* 2012;199:1241-6.
48. Sui X, Meinel FG, Song W, et al. Detection and size measurements of pulmonary nodules in ultra-low-dose CT

- with iterative reconstruction compared to low dose CT.
Eur J Radiol 2016;85:564-70.
49. Zhao YR, van Ooijen PM, Dorrius MD, et al. Comparison

of three software systems for semi-automatic volumetry of pulmonary nodules on baseline and follow-up CT examinations. Acta Radiol 2014;55:691-8.

Cite this article as: Han D, Heuvelmans MA, Oudkerk M. Volume versus diameter assessment of small pulmonary nodules in CT lung cancer screening. Transl Lung Cancer Res 2017;6(1):52-61. doi: 10.21037/tlcr.2017.01.05