Evaluation of Accuracy and Precision of Manual Size Measurements in Chest Tomosynthesis using Simulated Pulmonary Nodules

Christina Söderman, MSc, Åse A. Johnsson, MD, PhD, Jenny Vikgren, MD, PhD, Rauni R. Norrlund, MD, PhD, David Molnar, MD, Angelica Svalkvist, PhD, Lars G. Månsson, PhD, Magnus Båth, PhD

Rationale and Objectives: To investigate the accuracy and precision of pulmonary nodule size measurements on chest tomosynthesis images.

Materials and Methods: Artificial ellipsoid-shaped nodules with known sizes were inserted in clinical chest tomosynthesis images. The volume of the nodules corresponded to that of a sphere with a diameter of 4.0, 8.0, or 12.0 mm. Four thoracic radiologists were given the task to determine the longest diameter of the nodules. All nodules were measured twice. Measurement accuracy in terms of the mean measurement error was determined. Intraobserver and interobserver variabilities, as well as variability because of differences between nodules and their locations, were used as measures of precision.

Results: The mean measurement error ranged from −0.3 to 0.1 mm for the nodule size groups and observers. Of the smallest nodules, the observers found 7–17 of total 50 nodules nonmeasurable. The intraobserver and interobserver variabilities were of similar magnitude, indicating relatively small differences between the observers. The internodule variability was in general larger, indicating that the different characteristics of the nodules and their location are sources of variability.

Conclusions: The results suggest a high accuracy and precision for manual measurements of the nodules in chest tomosynthesis images. However, small nodules (<5.0 mm) may be difficult to measure at all because of poor visibility.

Key Words: Thoracic radiography; tomosynthesis; solitary pulmonary nodule; dimensional measurement accuracy; observer variation.

Pulmonary nodules are common incidental findings from chest and cardiac computed tomography (CT) examinations (1,2). As indicators of malignancy for these lesions, nodule size and growth are important factors (3,4). The Fleischner society has proposed a generally accepted management strategy for the follow-up of indeterminate nodules, which is based on repeated CT examinations at certain time intervals, depending on the size of the nodule when it is detected (5). If a significant size increase in the nodule over time is detected in the images, further diagnostic investigations are initiated. Similar recommendations can be found in the Response Evaluation Criteria In Solid Tumors (RECIST) guidelines concerning the evaluation of treatment response for tumors and metastases (6). The noninvasive nature of the technique and the high conspicuity of lesions in the images make CT scanning a suitable choice for follow-up. However, repeated CT scans raise concerns regarding the patient radiation dose burden (7).

Chest tomosynthesis has relatively recently emerged as an interesting alternative modality in lung imaging (8–11). The technique allows imaging of the chest in section images at a radiation dose substantially lower than in the case with chest CT (12,13). The financial cost for a chest tomosynthesis

Acad Radiol 2015; 22:496–504

From the Department of Radiation Physics, Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, SE-41345 Gothenburg, Sweden (C.S., L.G.M., M.B.); Department of Radiology, Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden (A.A.J., J.V., R.R.N.); Department of Radiology, Sahlgrenska University Hospital, Gothenburg, Sweden (A.A.J., J.V., R.R.N., D.M.); and Department of Medical Physics and Biomedical Engineering, Sahlgrenska University Hospital, Gothenburg, Sweden (A.S., L.G.M., M.B.). Received August 29, 2014; accepted November 29, 2014. This work was supported by grants from the Swedish Research Council [2011/488, 2013/3477], the Swedish Radiation Safety Authority [2008/2232, 2009/1689, 2010/4363, 2012/2021, 2013/2982], the Swedish Federal Government under the LUA/ALF agreement [ALFGBG-136281], and the Health and Medical Care Committee of the Region Västra Götaland [VGFOUREG-12046, VGFOUREG-27551, VGFOUREG-81341]. **Address correspondence to:** C.S. e-mail: christina.e.soderman@vgregion.se

http://dx.doi.org/10.1016/j.acra.2014.11.012

examination is also, in general, lower than that for a chest CT examination. Additionally, patient throughput is higher for the chest tomosynthesis examinations (11). Although the depth resolution of chest tomosynthesis is limited compared to CT, previous studies have shown that a relatively large amount of pulmonary nodules visible on CT images are detectable on chest tomosynthesis images (14–17). Furthermore, and of even higher importance for the present study, in the study by Vikgren et al. (15) all nodules with a longest diameter larger than 6 mm confirmed on CT were deemed visible in retrospect by tomosynthesis. This indicates the potential of using tomosynthesis for follow-up of already detected nodules.

Assessment of lesion size is commonly done by manual diameter measurements in the images. Inevitably, this will lead to observer variability, affecting the reliability of the measurement and the sensitivity of detecting any changes in the size. Previous studies have indicated good agreement between manual nodule measurements on CT and tomosynthesis (18,19). Limiting factors in these studies are that the measurements were done either on a nonanatomical background (18), restricting the clinical validity of the results, or on real nodules in clinical images where the true nodule size was unknown (19), making it difficult to establish any systematic errors in the measurements. To address these issues, the aim of the present study was to investigate the accuracy and precision of pulmonary nodule measurements on chest tomosynthesis images by using artificial nodules with known sizes located in real anatomical backgrounds.

MATERIALS AND METHODS

Image Collection

A subset of the patient images used in a previous study by Svalkvist et al. (20) was used as a basis for the creation of the image material in the present study. All included patients were referred for a combined chest CT and conventional chest radiography examination. For study purposes, a chest tomosynthesis examination was performed in addition to the referred examinations. This was approved by the Regional Ethical Review Board, and all participants gave written informed consent.

The tomosynthesis examinations were performed in conjunction with and on the same equipment as the conventional chest radiography examinations (Definium 8000 with VolumeRAD option; GE Healthcare, Chalfont St Giles, UK). With this system, a tomosynthesis examination consists of 60 low-dose projection images, acquired at a tube voltage of 120 kVp and within an angular interval of +15° to −15° relative to the standard orthogonal posteroanterior projection. The x-ray output used for all the projection images is determined by the resulting exposure of a scout view image, which is a conventional posteroanterior radiograph acquired with automatic exposure control before the tomosynthesis projection image acquisition. The tube load used for the scout view is multiplied by a user-adjustable dose ratio and distributed evenly between the 60 tomosynthesis projection images. In the present study, a dose ratio of 10:1 was used for all examinations. The resulting tube load per projection image is rounded down to the closest value possible on the system, with the restriction of a minimum possible tube load per projection of 0.25 mAs. During the image acquisition, which takes approximately 10 seconds, the x-ray tube moves continuously in the vertical direction while the position of the detector is fixed. Filtered backprojection is used to reconstruct coronal section images from the projection images. Normally, approximately 60 section images are reconstructed with a 5 mm interval. The pixel size in the projection radiographs and in the reconstructed images is 0.2×0.2 mm^2. The typical effective dose for a tomosynthesis examination on this system has been reported to be 0.13 mSv for a standard size patient (170 cm/70 kg) (13). Using the method for estimating the dose-area product of a chest tomosynthesis examination described by Båth et al. (21) and the proposed conversion factor of 0.26 mSv/Gy cm^2 (13), the average effective dose for the patients included in the present study was also determined to be 0.13 mSv.

Nodule Simulation

Artificial three-dimensional nodules, with a voxel size of $0.1 \times 0.1 \times 0.1$ mm^3, were simulated. The nodules were given an ellipsoid shape with equal length for the two minor axes. The volume of each nodule corresponded to that of a sphere with a diameter of 4.0, 8.0, or 12.0 mm. Each of the three size groups consisted of 50 nodules with a uniform distribution of the ratio between the lengths of the major and minor axes ranging from 1.1 to 1.5. This gave in total 150 nodules with a variety of shapes, from a sphere-like shape to a more elongated ellipsoid shape. The ranges of the length of the major axis were 4.3–5.2, 8.5–10.5, and 12.8–15.7 mm for the three nodule size groups.

Nodule Insertion

The simulated nodules were projected into the raw data projection images from the tomosynthesis examination of the included patients, using a method described previously in detail by Svalkvist et al. (20,22). Knowledge of the geometry of the tomosynthesis system is used in the method to convert a particular desired position of the nodule in the patient to a position in each of the 60 projection images. The paths of the x-rays from the focal spot to the detector are traced and the amount with which they are attenuated as they pass through the simulated nodule is determined. By adjusting the pixel values in the projection image raw data according to the resulting signal loss in the detector that would have occurred, had the nodule been present in the patient, the nodule is inserted in the images. Contrast loss because of scattered radiation and signal spread in the detector is accounted for in the insertion process. A certain amount of blurring caused by patient motion will always be present in tomosynthesis images. This is accounted for by shifting the nodule position, relative to the original position,

before each insertion of the nodule in each projection image. The direction of the shift in nodule position is randomized over all directions according to a uniform distribution. The amount of the shift is randomized according to a normal distribution with zero mean and a given standard deviation (SD_{motion}) (see subsequently).

All nodules were assigned a homogenous density and an attenuation coefficient of 0.16/cm. The value of the attenuation coefficient was chosen to represent a typical nodule density based on measurements made by Svalkvist et al. (20) of the CT number of real nodules. The attenuation in the existing anatomy at the location of the inserted nodule was compensated for by subtracting the attenuation coefficient of lung tissue from the nodule attenuation coefficient. In the present study, an attenuation coefficient of 0.03/cm was used for lung tissue, which was also determined by Svalkvist et al. (20) by CT number measurements.

The method for inserting simulated nodules in clinical images was validated by Svalkvist et al. (20) in a previous study, where 129 simulated nodules were inserted into patient images at positions in the lung that resembled those of real nodules found in patients. Fifty of these positions were selected for the simulated nodules in the present study. This resulted in 50 unique positions in 28 different chest tomosynthesis image sets from an equal number of examinations and patients. Each image set was used for between 1 and 4 simulated nodules. Nodules from the three different size groups were created to correspond to each other in the ratio of the length of the major and minor axes and were inserted at identical positions.

In the study by Svalkvist et al. (20), SD_{motion} was adjusted for each nodule so that the blurring because of patient motion of the nodules visually matched that of the surrounding anatomical structures at a specific location in the patient. The determined values for SD_{motion} were also used in the present study, and for the subset of locations used the mean value of SD_{motion} was 0.23 mm, ranging from 0.14 to 0.32 mm depending on the location in the patient.

The insertion of the simulated nodules in the tomosynthesis projection images was such that the ellipsoid major axis lies in the same plane as the final reconstructed section images. All nodules were centered in the depth direction in one of the section images. This was not the case in the study by Svalkvist et al., therefore small adjustments had to be made to the nodule positions in this direction. In some of the cases, adjustments in the two other directions had to be made as well to fit the entire volume of the simulated nodule within the lung parenchyma and not to position the nodule unnaturally close to other anatomical structures, such as the ribs or diaphragm. As the method does not take into account any morphologic changes of the surrounding anatomy because of the presence of nodule, the validity of the method is limited close to structures that would be affected by the presence of a nodule. A subjective evaluation of the resulting images indicated that this was not a problem in any of the cases in the present study. A rotation was introduced to all the nodules before they were inserted in the images, such that the major axis of the nodules lies in different directions in the image plane. The amount of rotation was chosen randomly but was equal for the nodules in each of the three size groups at the same location.

Measurement Study

Four thoracic radiologists participated in a measurement study where the task was to determine the longest diameter of the simulated nodules on the chest tomosynthesis images. Three of the observers had approximately 20 years of experience in thoracic radiology and 6–7 years of experience in chest tomosynthesis. The fourth observer had 1 year of experience in both thoracic radiology and chest tomosynthesis. The images were displayed with ViewDEX (23,24), a software tool intended for use in different types of observer studies with medical images, which includes the possibility of performing size measurements with electronic calipers in the images. The observers manually entered the result of each measurement directly in the software.

To study the intraobserver variability, 150 nodules were presented twice in a random order to the observers in two separate sessions. There were no instructions to the observers to wait for a certain time before continuing with the second session, because the risk of memory bias was considered small. The observers were aware that the nodules were artificial and ellipsoid in shape, but no other information regarding the shape or the size was given.

A subset of 10 consecutive section images from each patient around each nodule was presented to the observers. The nodules were in focus in one of the central section images. A circular region of interest (ROI) in the images was used to mark the position of the nodule. All nodules were centered in the ROI, which was present in all 10 section images and was of equal size for all nodules. In Figure 1, examples of nodules included in the study are shown. By default, the images were shown with a certain degree of magnification, but the observers were free to use the magnification tool additionally and to adjust the window width and level. The default settings for window width and level were set according to that specified by the digital imaging and communications in medicine (DICOM) header. Information about in which of the 10 section images the observers made each measurement was logged.

If the observers deemed the nodule not visible or not adequately reproduced for measurement, they had the alternative of stating the nodule as nonmeasurable.

Data Analysis

As a measure of the accuracy of the nodule size assessment, mean measurement error in the first session for each observer and for each of the three different nodule size groups was determined. Three measures of the precision of the nodule size measurements were used: 1) Repeated measurements on the same nodule by a given observer will always be associated with a certain variation. For each observer, the standard

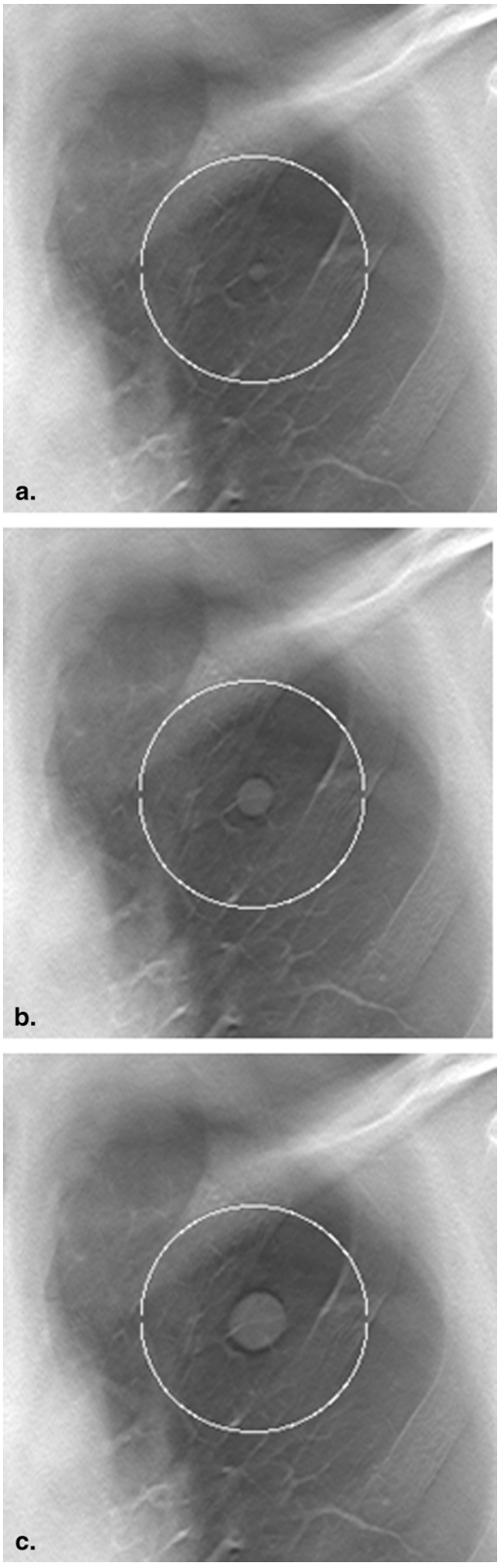

Figure 1. Cutouts from reconstructed chest tomosynthesis section images included in the measurement study showing an example of simulated nodules from the three different size groups corresponding to each other in terms of the ratio between the lengths of the major and minor axes, amount of applied rotation, and location in the lung. The longest diameters of the nodules are **(a)** 4.3 mm, **(b)** 8.6 mm, and **(c)** 12.9 mm.

deviation of the two measurements on each nodule, averaged over all nodules in each size group, was used to express this variation and referred to as the intraobserver variability, 2) Measurements on the same nodule by multiple observers will lead to an increased variation because of differences between observers. The standard deviation of the measurements by all observers on each nodule, averaged over all nodules in each size group, was used to express this variation and referred to as interobserver variability, and 3) The different characteristics of the nodules and the anatomy at their locations will act as an additional source of measurement variation. For each observer, the standard deviation of the measurement error for all nodules in each size group was used to express this variation and referred to as the internodule variability. The three variability measures are not independent, as both the interobserver and internodule variabilities will include the intraobserver variability. A correction for small sample sizes was used in the determination of the standard deviations on which the intraobserver (sample size 2) and interobserver (sample size 4) variablities were based (25). The intraobserver and interobserver variabilities were also assessed with the intraclass correlation coefficient. The 95% confidence interval for each measure was used to express the uncertainty of the results. The proportion of nodules judged as nonmeasurable by each observer for the different nodule size groups was determined.

Any measurement that was judged to have been done on a structure other than a simulated nodule was excluded from the analysis of the measurement accuracy and precision. This decision was based on a visual inspection of the images in combination with the log of in which of the section images the measurements were done.

RESULTS

The maximum number of excluded measurements for any of the observers was five (observer 4). All excluded measurements except one were made in images containing nodules from the smallest size group.

The number of nodules judged as nonmeasurable, the mean absolute measurement error, and the mean relative measurement error for each observer and nodule size group are presented in Table 1. Mean values are presented with 95% confidence intervals.

Of the smallest nodules, the observers found 7–17 of total 50 nodules nonmeasurable. For the nodules corresponding in volume to a sphere with a diameter of 8 mm, this number ranged from 0 to 3. All observers found all the large nodules measurable.

Similar mean measurement errors ranging from −0.3 to 0.1 mm were found for all observers and nodule sizes. One of the observers (observer 4) significantly underestimated the size of the longest diameter of the smallest nodules. For observers 1 and 3, the results indicated a limited underestimation for all nodule sizes. For observer 2, no indication of overestimation or underestimation of the longest diameter of any of the nodule sizes was found.

TABLE 1. Number of Nodules Judged as Nonmeasurable, Mean Measurement Error, and Mean Relative Measurement Error for Each Observer for the Three Different Size Groups

Parameter	Observer 1	Observer 2	Observer 3	Observer 4
4.0 mm				
N nonmeasurable	15	16	17	7
Mean error (mm)	−0.2 (−0.3, 0.0)	0.0 (−0.1, 0.2)	−0.2 (−0.3, 0.0)	−0.3 (−0.5, −0.2)
Mean relative error (%)	−3.3 (−6.5, 0.0)	0.5 (−2.7, 3.7)	−3.3 (−6.2, −0.5)	−7.2 (−10.4, −4.0)
8.0 mm				
N nonmeasurable	1	3	2	—
Mean error (mm)	−0.2 (−0.4, 0.0)	0.1 (−0.1, 0.2)	−0.2 (−0.4, 0.0)	−0.2 (−0.4, 0.0)
Mean relative error (%)	−1.8 (−3.9, 0.2)	0.6 (−0.8, 2.0)	−2.3 (−4.4, −0.1)	−2.0 (−3.8, −0.2)
12.0 mm				
N nonmeasurable	—	—	—	—
Mean error (mm)	−0.1 (−0.2, 0.0)	−0.1 (−0.2, 0.1)	−0.1 (−0.2, 0.0)	0.1 (0.0, 0.2)
Mean relative error (%)	−0.5 (−1.2, 0.1)	−0.5 (−1.4, 0.5)	−0.7 (−1.6, 0.2)	0.6 (−0.2, 1.5)

Data in parentheses are 95% confidence intervals.

Figure 2 shows the measurement error for each of the 50 nodules in the three size groups for all observers. It can be noted that the longest diameter of a number of nodules was noticeably underestimated. By visual inspection it was found that most of these nodules were located in regions corresponding to low dose regions in the projection images, for example, behind the heart or the diaphragm. For the nodules corresponding in volume to a sphere with a diameter of 8.0 or 12.0 mm, this was the case for all measurements for which the underestimation was 1.5 mm or larger. Nodules in the smallest size group with corresponding locations were all deemed not measurable by all observers.

The intraobserver, interobserver, and internodule variabilities are presented in Tables 2–4, respectively. The intraobserver and interobserver variabilities were of similar magnitude, indicating relatively small differences between observers. However, the internodule variability was in general larger, indicating that the different characteristics of the nodules and their location are sources of variability. In terms of absolute measurements, all measures of variability showed little or no dependency on the nodule size. However, for all observers there was a tendency of decreasing intraobserver and internodule variabilities with nodule size in terms of the relative measurements. A similar tendency was found for the interobserver variability. The intraclass correlation coefficient was in general high, ranging from 0.71 to 0.99 for the intraobserver variability and from 0.91 to 0.97 for the interobserver variability.

DISCUSSION

Chest tomosynthesis is a relatively new technique allowing imaging of the chest in coronal section images. It has been shown to improve the visibility and detection rate of lung nodules compared to conventional radiography (14,15). However, the role of tomosynthesis in chest radiology has not yet been fully established. One clinical task where the technique might be beneficial to use is in the follow-up of known lung nodules, which has been proposed previously by Dobbins and McAdams (8). The objective of the present study was to evaluate the accuracy and precision of nodule size measurements on chest tomosynthesis images. Therefore, simulated nodules were used because this enables the possibility of knowing the true sizes of the nodules exactly. The artificial nodules were inserted into patient chest tomosynthesis images, and the longest diameter of the nodules was determined by four thoracic radiologists.

Because the longest diameter in the plane of detected nodules can be found in any direction, ellipsoid-shaped nodules were created and inserted in the images so that the major axis lies in different directions in the image plane. There was no detection task involved in the study, enabling the position of the nodules to be marked with an ROI. By not informing the observers in which of the section images presented to them the nodule was in focus, the uncertainty of not knowing where the nodule is at its largest was included in the study.

By positioning the simulated nodules so that they were centered in the depth direction in one of the section images, an ideal condition for assessing the nodule size was created in the present study. In the clinical situation, the uncertainty in matching the nodule location with the plane of reconstruction can be expected to affect the accuracy in determining the longest nodule diameter. There is, however, a possibility of reconstructing section images at smaller intervals than was done in the present study and thereby increasing the probability of including the longest diameter of the nodules in the images.

In a previous similar phantom study (18), where measurements were done on spheres of various sizes and attenuation coefficients in a nonanatomical background, absolute and relative mean measurement errors for all spheres and observers were −0.10 mm and −1.1%, respectively, for chest tomosynthesis. The corresponding values for CT were 0.15 mm and 1.4%. The measurement accuracy in the present study was comparable to that found for chest tomosynthesis and CT in the phantom study. Regarding measurement precision, in

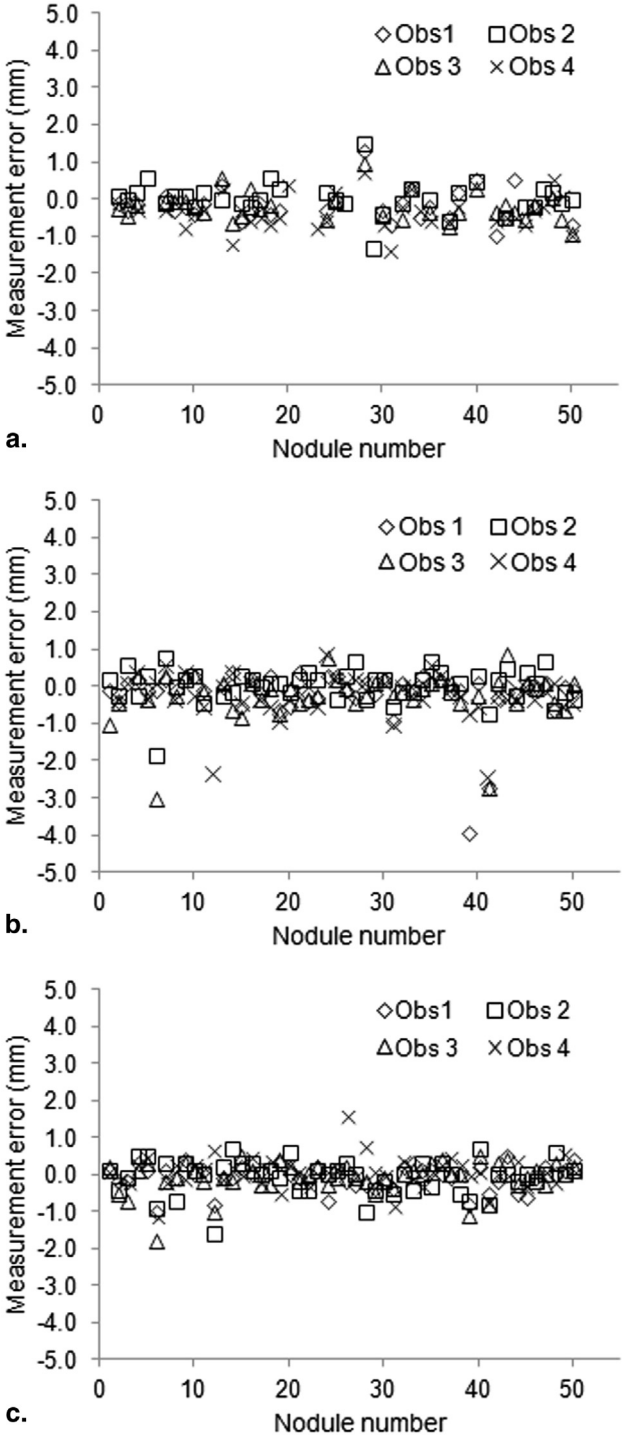

Figure 2. Measurement error for all observers for nodules corresponding in volume to a sphere with a diameter of **(a)** 4 mm, **(b)** 8 mm, and **(c)** 12 mm. Nodules with the same number in the three different nodule size groups were located at identical positions in the images.

the phantom study, the standard deviations of the measurement error were 0.3 mm and 5.0% for chest tomosynthesis. In the present study, the internodule variability, for all nodule sizes and observers, ranged from 0.3 to 0.7 mm and 2.3% to 9.5%, indicating a somewhat lower measurement precision in the present study. This is expected, given the more realistic variation in nodule shape and image background used in the present study. The largest absolute internodule variability was observed for the nodules corresponding in volume to a sphere with a diameter of 8 mm. This can in large part be attributed to the nodules in this size group for which the longest diameter was largely underestimated because of difficulties in distinguishing the contours of the nodules in low dose areas of the image. In the previous phantom study, the task of the observers was to measure the left-to-right diameter of the spheres. The possibly more difficult—but on the same time more clinically relevant—task in the present study of determining the longest diameter of the nodules may also have contributed to a larger spread in the data.

In another similar study by Johnsson et al. (19), the right-to-left and inferior-to-superior diameters of real clinical nodules were measured in chest tomosynthesis images. Measurement error was evaluated relative to segmented diameters on CT. Reported standard deviations of the measurement errors of the observers ranged from 2.2 to 2.6 mm and from 2.0 to 2.3 mm for the left-to-right and inferior-to-superior diameters, respectively. These values are substantially larger than the internodule variabilities found in the present study. Pulmonary nodules found in patients are of varying shapes and with irregular surfaces. This will inevitably affect the variability of nodule size measurements in a manner that was not accounted for in the present study. This may partly explain the larger internodule variability presented by Johnsson et al. compared to the present study. However, it should be noted that the true nodule size was not known in the study by Johnsson et al., which may have influenced the validity of the reported measurement variation as the accuracy of the segmentation process affected the result. In the present work, the true nodule size was known exactly.

In the case of CT imaging, it has previously been shown that the influence of window setting on the size assessment of pulmonary nodules can be large (26). In the present study, the observers were free to change the window settings to find the best condition in which to measure the nodule, possibly affecting the presented measurement variability. However, the effect of window setting on apparent nodule size can be expected to be smaller for tomosynthesis than for CT because of the better in-plane resolution of tomosynthesis, as pointed out by Johnsson et al. (18). Furthermore, the observers did not use the possibility to change the window settings to a large extent. One observer did all measurements in the default window settings, whereas the other three observers altered the window settings in approximately 40%, 20%, and 10% of the cases, respectively.

The choice of follow-up strategy of incidentally detected pulmonary nodules, as proposed by the Fleischner society, is dependent on the size of the nodule when it is detected, where the categorization of nodule size is divided into the four groups $\leq$ 4 mm, $>$ 4-6 mm, $>$ 6-8 mm and $>$ 8 mm. Considering the overall measurement accuracy and precision found in the present study, the risk of assigning a nodule to an

TABLE 2. Intraobserver Variability Expressed as the Standard Deviation of the Two Measurements on Each Nodule, Both in Absolute Terms and Relative to the Actual Size of the Nodules, Averaged Over all Nodules in Each Size Group for Each Observer

Parameter	Observer 1	Observer 2	Observer 3	Observer 4
4.0 mm				
Intraobserver variability (mm)	0.2 (0.1, 0.2)	0.4 (0.3, 0.5)	0.2 (0.2, 0.3)	0.3 (0.2, 0.4)
Intraobserver variability (%)	3.6 (2.3, 5.0)	7.7 (5.4, 9.9)	5.3 (3.5, 7.1)	6.1 (4.4, 7.8)
Intraclass correlation coefficient	0.92 (0.83, 0.96)	0.71 (0.32, 0.87)	0.83 (0.57, 0.93)	0.81 (0.62, 0.90)
8.0 mm				
Intraobserver variability (mm)	0.2 (0.1, 0.3)	0.3 (0.2, 0.3)	0.2 (0.2, 0.3)	0.3 (0.2, 0.4)
Intraobserver variability (%)	2.0 (1.4, 2.7)	2.8 (2.1, 3.4)	2.4 (1.7, 3.1)	3.0 (2.3, 3.8)
Intraclass correlation coefficient	0.94 (0.89, 0.97)	0.91 (0.84, 0.95)	0.95 (0.90, 0.97)	0.92 (0.86, 0.96)
12.0 mm				
Intraobserver variability (mm)	0.2 (0.1, 0.2)	0.3 (0.3, 0.4)	0.2 (0.2, 0.3)	0.4 (0.2, 0.5)
Intraobserver variability (%)	1.1 (0.9, 1.4)	2.4 (1.9, 2.9)	1.5 (1.1, 1.8)	2.7 (1.7, 3.7)
Intraclass correlation coefficient	0.99 (0.98, 0.99)	0.94 (0.89, 0.96)	0.98 (0.96, 0.99)	0.87 (0.78, 0.93)

Data in parentheses are 95% confidence intervals.

TABLE 3. Interobserver Variability Expressed as the Standard Deviation of the Measurements, Both in Absolute Terms and Relative to the Actual Size of the Nodules, by all Observers on Each Nodule, Averaged Over all Nodules in Each Size Group

Parameter	4 mm	8 mm	12 mm
Interobserver variability (mm)	0.2 (0.2, 0.3)	0.3 (0.2, 0.4)	0.3 (0.2, 0.4)
Interobserver variability (%)	5.3 (4.1, 6.5)	3.3 (2.4, 4.2)	2.1 (1.7, 2.5)
Intraclass correlation coefficient	0.91 (0.82, 0.96)	0.93 (0.88, 0.96)	0.97 (0.96, 0.98)

Data in parentheses are 95% confidence intervals.

TABLE 4. Internodule Variability, Expressed as the Standard Deviation of the Measurement Error, Both in Absolute Terms and Relative to the Actual Size of the Nodule, for all Nodules in Each Size Group for Each Observer

Parameter	Observer 1	Observer 2	Observer 3	Observer 4
4.0 mm				
Internodule variability (mm)	0.4 (0.3, 0.5)	0.4 (0.3, 0.6)	0.4 (0.3, 0.5)	0.5 (0.4, 0.6)
Internodule variability (%)	9.0 (7.2, 11.9)	9.0 (7.2, 11.9)	7.9 (6.3, 10.5)	9.5 (7.8, 12.4)
8.0 mm				
Internodule variability (mm)	0.7 (0.6, 0.9)	0.4 (0.4, 0.6)	0.7 (0.6, 0.8)	0.6 (0.5, 0.7)
Internodule variability (%)	7.2 (6.0, 9.0)	4.8 (4.0, 6.0)	7.2 (6.0, 9.1)	6.2 (5.2, 7.8)
12.0 mm				
Internodule variability (mm)	0.3 (0.3, 0.4)	0.5 (0.4, 0.6)	0.4 (0.4, 0.5)	0.4 (0.4, 0.5)
Internodule variability (%)	2.3 (1.9, 2.9)	3.3 (2.7, 4.1)	3.1 (2.6, 3.9)	3.0 (2.5, 3.7)

Data in parentheses are 95% confidence intervals.

inappropriate follow-up strategy is small with the use of chest tomosynthesis. However, in the case of follow-up, the repeated imaging will act as an additional source of variability not investigated in the present study. Differences in, for example, the patient positioning between two tomosynthesis scans may result in that the nodule will be located differently in relation to the section image reconstruction geometry. The final section images will in turn match the location of the nodule differently, possibly leading to differences in the perceived size of the nodule between two time points. Further investigations are necessary to determine how this variability will affect the possibility of detecting changes in the nodule size in the coronal plane by chest tomosynthesis.

Compared to the other two nodule size groups, a relatively large proportion of nodules in the smallest size group was judged as not measurable by the observers (Table 1). This somewhat contradicts the finding in the nodule detection study by Vikgren et al. (15), where 86% (44 of 51) of lung nodules with a longest diameter of 4.0 mm or smaller found on CT images were visible on chest tomosynthesis images. However, the reference method in the detection study was based on CT images with a slice thickness of 5 mm, possibly causing small nodules not to be found by the reference and thereby not included in the study. In comparison, Asplund et al. (27), in a similar study but where the reference was based on CT images with a slice thickness of 0.6 or 1.25 mm, found

that 52% of nodules with a longest diameter of 4.0 mm or smaller were found visible on chest tomosynthesis images. In the previously mentioned phantom measurement study (18), some of the smallest spheres were judged as nonmeasurable on both CT and chest tomosynthesis images. Dobbins and McAdams (8) have previously expressed the risk of missing new smaller nodules in the case of chest tomosynthesis being used as lung nodule follow-up modality.

Although the number of the smallest nodules deemed as not measurable was relatively large, the measurement accuracy and precision for the measured nodules in this size group were high. These results may indicate that the observers chose to only measure the nodules for which the borders could be identified above a certain degree of confidence. Had the observers been forced to measure all nodules, one could suspect that the accuracy and precision would have been less conclusive. In the present study, the observer with the smallest number of nonmeasurable nodules (observer 4) was the only observer who significantly underestimated the nodule size.

In Figure 1, a dark halo-like artifact can be seen around the inserted nodules. This is a well-known artifact found in reconstructed tomosynthesis section images and is a result of the incomplete sampling of frequency space, which in turn is because of the limited angular coverage by the projection images. The artifact will appear in the same direction as that in which the tomosynthesis scan is performed. In the case of chest tomosynthesis images, the artifact will appear in the craniocaudal direction. The present study was not designed to address how this effect influences pulmonary nodule measurements. It could be beneficial to further investigate this to evaluate how the artifact should be managed in nodule size assessment to achieve the best accuracy, in particular when the longest diameter is parallel to the scan direction. A preliminary analysis of the measurement data, in conjunction with the fact that the observers stated that they did not include the artifact when measuring the nodules, indicates that the halo should not be included when measuring nodules.

A number of limiting factors were present in the study. Firstly, the observers were aware that the nodules had an ellipsoid shape. This might have helped them to guess the outline of the nodule in any cases where the nodule was obscured by other anatomical structures. Secondly, the fact that the nodules were centered in a circular ROI in the images may also have affected the observers, such that the borders of the ROI could be used as a helping reference when determining the diameter of the nodule. However, this effect was considered to be limited as the ROI was relatively large compared to the size of the nodules. Additionally, the observers were not explicitly informed that the nodules were centered in the ROI. (As described previously, some measurements had to be excluded because it was evident that they were made on a structure other than a nodule.) Thirdly, and as mentioned previously, the present study does not take into account the loss of measurement accuracy and precision because of irregular nodule borders or mismatch between the nodule location and image reconstruction plane. Furthermore, the present study did not investigate the possibility of detecting changes in the nodule size over time in chest tomosynthesis images. This needs to be analyzed before chest tomosynthesis may be used for follow-up of pulmonary nodules.

In conclusion, the results of the present study suggest a high accuracy and precision for manual measurements of nodules in chest tomosynthesis images, although a limited underestimation of the size is indicated. Small nodules (<5.0 mm) may, however, be difficult to measure at all because of poor visibility.

REFERENCES

1. Burt JR, Iribarren C, Fair JM, et al. Incidental findings on cardiac multidetector row computed tomography among healthy older adults: prevalence and clinical correlates. Arch Intern Med 2008; 168:756–761.
2. Hall WB, Truitt SG, Scheunemann LP, et al. The prevalence of clinically relevant incidental findings on chest computed tomographic angiograms ordered to diagnose pulmonary embolism. Arch Intern Med 2009; 169: 1961–1965.
3. Hasegawa M, Sone S, Takashima S, et al. Growth rate of small lung cancers detected on mass CT screening. Br J Radiol 2000; 73:1252–1259.
4. Henschke CI, Yankelevitz DF, Naidich DP, et al. CT screening for lung cancer: suspiciousness of nodules according to size on baseline scans. Radiology 2004; 231:164–168.
5. 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.
6. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer 2009; 45:228–247.
7. Brenner DJ, Hall EJ. Computed tomography—an increasing source of radiation exposure. N Engl J Med 2007; 357:2277–2284.
8. Dobbins JT, III, McAdams HP. Chest tomosynthesis: technical principles and clinical update. Eur J Radiol 2009; 72:244–251.
9. Johnsson ÅA, Vikgren J, Svalkvist A, et al. Overview of two years of clinical experience of chest tomosynthesis at Sahlgrenska University Hospital. Radiat Prot Dosimetry 2010; 139:124–129.
10. Quaia E, Baratella E, Cioffi V, et al. The value of digital tomosynthesis in the diagnosis of suspected pulmonary lesions on chest radiography: analysis of diagnostic accuracy and confidence. Acad Radiol 2010; 17: 1267–1274.
11. Johnsson ÅA, Vikgren J, Båth M. Chest tomosynthesis: technical and clinical perspectives. Semin Respir Crit Care Med 2014; 35:17–26.
12. Sabol JM. A Monte Carlo estimation of effective dose in chest tomosynthesis. Med Phys 2009; 36:5480–5487.
13. Båth M, Svalkvist A, von Wrangel A, et al. Effective dose to patients from chest examinations with tomosynthesis. Radiat Prot Dosimetry 2010; 139:153–158.
14. Dobbins TJ, McAdams HP, Song JW, et al. Digital tomosynthesis of the chest for lung nodule detection: interim sensitivity results from an ongoing NIH-sponsored trial. Med Phys 2008; 35:2554–2557.
15. Vikgren J, Zachrisson S, Svalkvist A, et al. Comparison of chest tomosynthesis and chest radiography for detection of pulmonary nodules: human observer study of clinical cases. Radiology 2008; 249: 1034–1041.
16. Zachrisson S, Vikgren J, Svalkvist A, et al. Effect of clinical experience of chest tomosynthesis on detection of pulmonary nodules. Acta Radiol 2009; 50:884–891.
17. Asplund S, Johnsson ÅA, Vikgren J, et al. Learning aspects and potential pitfalls regarding detection of pulmonary nodules in chest tomosynthesis and proposed related quality criteria. Acta Radiol 2011; 52: 503–512.
18. Johnsson ÅA, Svalkvist A, Vikgren J, et al. A phantom study of nodule size evaluation with chest tomosynthesis and computed tomography. Radiat Prot Dosimetry 2010; 139:140–143.
19. Johnsson ÅA, Fagman E, Vikgren J, et al. Pulmonary nodule size evaluation with chest tomosynthesis. Radiology 2012; 265:273–282.

20. Svalkvist A, Johnsson ÅA, Vikgren J, et al. Evaluation of an improved method of simulating lung nodules in chest tomosynthesis. Acta Radiol 2012; 53:874–884.
21. Båth M, Söderman C, Svalkvist A. A simple method to retrospectively estimate patient dose-area product for chest tomosynthesis examinations performed using VolumeRAD. Med Phys 2014; 41:101905.
22. Svalkvist A, Håkansson M, Ullman G, et al. Simulation of lung nodules in chest tomosynthesis. Radiat Prot Dosimetry 2010; 139:130–139.
23. Börjesson S, Håkansson M, Båth M, et al. A software tool for increased efficiency in observer performance studies in radiology. Radiat Prot Dosimetry 2005; 114:45–52.
24. Håkansson M, Svensson S, Zachrisson S, et al. ViewDEX: an efficient and easy-to-use software for observer performance studies. Radiat Prot Dosimetry 2010; 139:42–51.
25. Bolch BW. More on unbiased estimation of the standard deviation. Am Stat 1968; 22:27.
26. Harris KM, Adams H, Lloyd DC, et al. The effect on apparent size of simulated pulmonary nodules of using three standard CT window settings. Clin Radiol 1993; 47:241–244.
27. Asplund SA, Johnsson ÅA, Vikgren J, et al. Effect of radiation dose level on the detectability of pulmonary nodules in chest tomosynthesis. Eur Radiol 2014; 24:1529–1536.