


Pulmonary subsolid nodules: value of semi-automatic measurement in diagnostic accuracy, diagnostic reproducibility and nodule classification agreement

Hyungjin Kim^{1,2} · Chang Min Park^{1,2,3}  · Eui Jin Hwang^{1,2} · Su Yeon Ahn^{1,2} · Jin Mo Goo^{1,2,3}

Received: 12 July 2017 / Revised: 30 September 2017 / Accepted: 3 November 2017 / Published online: 1 December 2017
© European Society of Radiology 2017

Abstract

Objectives We hypothesized that semi-automatic diameter measurements would improve the accuracy and reproducibility in discriminating preinvasive lesions and minimally invasive adenocarcinomas from invasive pulmonary adenocarcinomas appearing as subsolid nodules (SSNs) and increase the reproducibility in classifying SSNs.

Methods Two readers independently performed semi-automatic and manual measurements of the diameters of 102 SSNs and their solid portions. Diagnostic performance in predicting invasive adenocarcinoma based on diameters was tested using logistic regression analysis with subsequent receiver operating characteristic curves. Inter- and intrareader reproducibilities of diagnosis and SSN classification according to Fleischner's guidelines were investigated for each measurement method using Cohen's κ statistics.

Results Semi-automatic effective diameter measurements were superior to manual average diameters for the diagnosis of invasive adenocarcinoma (AUC, 0.905–0.923 for semi-automatic measurement and 0.833–0.864 for manual measurement; $p < 0.05$). Reproducibility of diagnosis between the readers also improved with semi-automatic measurement

($\kappa = 0.924$ for semi-automatic measurement and 0.690 for manual measurement, $p = 0.012$). Inter-reader SSN classification reproducibility was significantly higher with semi-automatic measurement ($\kappa = 0.861$ for semi-automatic measurement and 0.683 for manual measurement, $p = 0.022$).

Conclusions Semi-automatic effective diameter measurement offers an opportunity to improve diagnostic accuracy and reproducibility as well as the classification reproducibility of SSNs.

Key Points

- Semi-automatic effective diameter measurement improves the diagnostic accuracy for pulmonary subsolid nodules.
- Semi-automatic measurement increases the inter-reader agreement on the diagnosis for subsolid nodules.
- Semi-automatic measurement augments the inter-reader reproducibility for the classification of subsolid nodules.

Keywords Carcinoma, non-small-cell lung · Multidetector computed tomography · Diagnosis, computer-assisted · Dimensional measurement accuracy · Observer variation

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00330-017-5171-7>) contains supplementary material, which is available to authorized users.

✉ Chang Min Park
cmpark.morphius@gmail.com

¹ Department of Radiology, Seoul National University College of Medicine, 101, Daehak-ro, Jongno-gu, Seoul 03080, Korea

² Institute of Radiation Medicine, Seoul National University Medical Research Center, Seoul, Korea

³ Seoul National University Cancer Research Institute, Seoul, Korea

Abbreviations

AIS	Adenocarcinoma in-situ
AUC	Area under the curve
CI	Confidence interval
CTDI _{vol}	Volume CT dose index
DLP	Dose-length product
D _{solid}	Diameter of solid portion
D _{SSN}	Diameter of subsolid nodule
HU	Hounsfield unit
IPA	Invasive pulmonary adenocarcinoma
Lung-RADS	Lung CT Screening Reporting and Data System
MIA	Minimally invasive adenocarcinomas

P_{solid}	Solid proportion within a nodule
R_{diff}	Percentage relative difference
ROC	Receiver operating characteristic curve
SSDE	Size-specific dose estimate
SSN	Subsolid nodule

Introduction

Pulmonary subsolid nodules (SSNs) represent a histological spectrum of entities from preinvasive lesions to invasive pulmonary adenocarcinomas (IPAs) [1]. However, the management strategy can significantly differ as preinvasive lesions can be followed up conservatively with CT scans or treated with limited resection without lymph node dissection [2, 3], whereas IPAs require surgical resection for which the standard procedure is lobectomy with lymph node dissection [4]. Therefore, discrimination of preinvasive lesions from IPAs is critical in the evaluation of SSNs. The prognosis of patients is also substantially distinct between preinvasive lesions and IPAs, and it is well known that patients with adenocarcinomas in-situ (AIS) or minimally invasive adenocarcinomas (MIAs) who undergo complete surgical resection should have 100 % or near 100 % 5-year disease-free survival [2].

The differentiation between preinvasive lesions and IPAs is currently based on their CT morphological features, sizes and clinical history [3, 5–10]. Quantitative imaging features such as mass, kurtosis, entropy and volume or mass doubling time can also be used for the differentiation of SSNs [5–7]. However, the most fundamental and practical method of differentiation would be the diameter measurement of nodules as well as its internal solid components. Specifically, previous studies have suggested size criteria for the diagnosis of malignant SSNs as 8–10 mm [3, 8, 9]. In current everyday practice, diameter measurements are performed manually using electronic calipers on the axial plane of a CT scan. Alternatively, however, semi-automatic segmentation software can be applied to calculate the nodule size. The advantages of semi-automatic measurement are that it is more accurate and reproducible [11, 12]. In addition, semi-automatic measurement is basically three-dimensional and not performed on an axial plane. Thus, it is potentially closer to the ground-truth of a tumour's dimensions. Past publications that have investigated semi-automatic measurements have focused on measurement accuracy and variability. However, the actual merit of semi-automatic measurement in the diagnostic process of SSNs has not yet been analysed to our knowledge. It is unclear whether the diagnostic performance for SSNs could be improved when semi-automatic measurements are utilized instead of the conventional manual measurements.

Thus, in this study, we hypothesized that semi-automatic diameter measurements would improve the accuracy and reproducibility in discriminating preinvasive lesions and MIAs

from IPAs appearing as SSNs and increase the reproducibility in classifying SSNs. We then performed manual and semi-automated measurements using a commercial software program and compared the datasets to substantiate our assumption.

Materials and methods

This retrospective analysis was approved by the Institutional Review Board of Seoul National University Hospital and written informed consent was waived.

Study population

Our study population comprised 104 consecutive patients who underwent preoperative non-enhanced CT using a SSN evaluation protocol and subsequent surgical resection between November 2014 and July 2016. Among the 104 patients, three patients whose pathological diagnoses did not belong to the adenocarcinoma spectrum including precursors (atypical adenomatous hyperplasia (AAH), AIS, MIA and IPA) were excluded. Seven additional patients were excluded owing to semi-automatic segmentation failure. In addition, five patients in whom nodules were classified as solid by at least one reader (E.J.H. and H.K. with 6 and 7 years of experience in chest CT, respectively), were also excluded. Consequently, 89 patients (39 men and 50 women; mean age \pm standard deviation, 61.9 \pm 9.7 years for all individuals, 64.3 \pm 8.5 years for men, and 60.1 \pm 10.3 years for women; range 45–77 years for men and 34–79 years for women; independent t-test, $p=0.044$) with 102 SSNs were included in our study (Fig. 1).

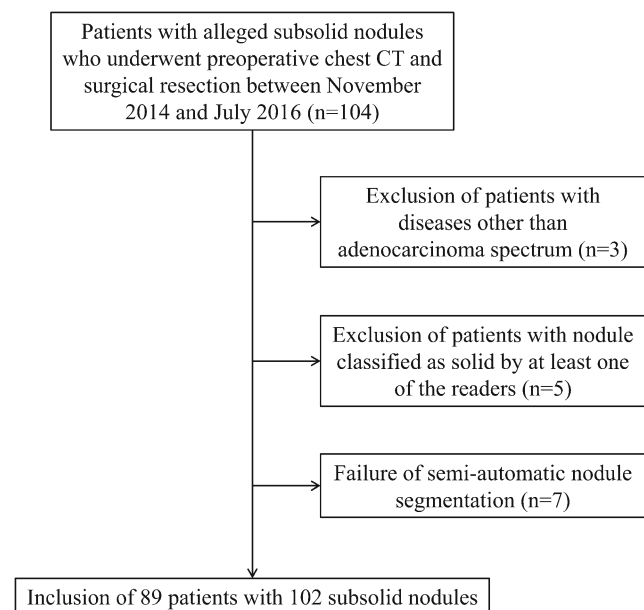


Fig. 1 Flow chart of the patient selection process including the inclusion and exclusion criteria

Seventy-eight of the patients had a single SSN; nine patients had two SSNs; and two patients had three nodules. Pathology revealed that there were two cases of AAH, three AIS, 27 MIAs and 70 IPAs. The median interval between CT and surgical resection was 2.0 ± 5.4 days (range 0–36 days). Part of the study population (36/89) had participated in a previous published study [13].

CT image acquisition

All patients were scanned in the supine position during maximum inspiration. Non-enhanced CT scans were performed with a 64-row multi-detector CT scanner (Discovery CT750 HD; GE Healthcare, Waukesha, WI, USA). Detailed scanning parameters were as follows: detector configuration, 64×0.625 mm; tube voltage, 120 kVp; noise index setting, 70.44; beam pitch, 0.984:1; rotation time, 0.5 s; field-of-view, 350 mm; matrix, 512×512 ; bone kernel; filtered back projection; reconstruction increment, 0.625 mm; and slice thickness, 0.625 mm.

For radiation dose assessment, the volume CT dose index ($CTDI_{vol}$) and dose-length product (DLP) were recorded from the scanner. The effective diameter, defined as the square root of the anteroposterior diameter times the transverse diameter [14], was obtained for each patient. Size-specific dose estimates (SSDE) were calculated to provide an approximation of the absorbed dose [14].

Semi-automatic and manual measurement of SSNs

First, semi-automatic measurement of the diameter of SSNs (D_{SSN}) and internal solid portions (D_{solid}) were performed by two radiologists (E.J.H. and H.K.) independently, using commercial software (Veolity version 1.2, MeVis Medical Solutions, Bremen, Germany) [13, 15, 16]. Readers initially determined and clicked one of the three nodule types (pure ground-glass, part-solid or solid) based on visual assessment according to Fleischner Society recommendations [2]. Thereafter, users defined a target by dragging a line traversing the nodule on an axial plane. Based on the predefined attenuation threshold for the ground-glass [-750 Hounsfield units (HU)] and solid portion (-350 HU) [15], the software determined the segmentation boundary [17]. Readers were allowed to adjust the segmentation boundary by tuning attenuation thresholds or the roundness of each component (ground-glass and/or solid portion), if necessary. Successful segmentation was evaluated subjectively by each reader (Electronic Supplementary Material 1) and was defined as correct segmentation proportion of 80 % or greater according to previous literature [13, 15, 18]. After segmentation, the effective diameters (diameter of a sphere where the volume equals the nodule volume) of both D_{SSN} and D_{solid} were provided. Thus, the

nodule type, D_{SSN} and D_{solid} assessed by both readers were recorded for the successfully segmented SSNs.

Second, manual measurements were recorded using the longest diameter with its maximum perpendicular diameter so as to obtain the average diameter of SSNs (D_{SSN}) and their solid portions (D_{solid}). The radiologists were asked to classify the nodule type and then to measure the diameters on the lung window (window width and level of 1,500 and -700 HU) using the electronic caliper of the picture archiving and communication system.

Finally, one of the readers (H.K.) conducted the overall image analysis twice to perform intrareader variability analysis. Therefore, we obtained three sets of measurement data (reader 1, reader 2–1 and reader 2–2). Each image reading session was separated by an interval of 4 weeks and the readers were unaware of the pathological diagnosis. The order of image reading was randomized by patient.

Pathological diagnosis

In our institution, all surgical specimens containing subsolid nodules were fixed by infusing 10 % buffered formalin through a transpleural and transbronchial approach to produce an inflated state and to precisely measure the invasive adenocarcinoma component [19]. Haematoxylin and eosin were used for staining the specimens. All pathological diagnoses were established based on the surgical specimens by attending pulmonary pathologists of Seoul National University Hospital according to the 2011 International Association for the Study of Lung Cancer (IASLC)/the American Thoracic Society (ATS)/the European Respiratory Society (ERS) pulmonary adenocarcinoma classification [20]. Pathological diagnoses were made after operations as a part of routine clinical process and specimens were not reviewed again, particularly for the present study.

Statistical analysis

Semi-automatic and manual measurements of D_{SSN} and D_{solid} were compared using either the paired t-test or Wilcoxon's test, as appropriate, after testing for normality. Thereafter, inter- and intrareader measurement variability of D_{SSN} and D_{solid} was calculated using the Bland-Altman method by calculating the 95 % confidence interval (CI) of the relative differences (differences in the two measurements divided by the mean) [21]. Percentage relative differences (R_{diff}) were also compared between the semi-automatic and manual measurements.

Diagnostic performance based on the diameter measurements was tested using binary logistic regression analysis with subsequent receiver operating characteristic curve (ROC) analysis. First, D_{SSN} and the solid proportion (%) within a nodule (P_{solid}) were used as input variables for logistic

regression analysis using the pathological diagnosis (preinvasive lesion (AAH, AIS) and MIA vs. IPA) as a dependent variable in order to evaluate the statistical significance of the two variables. After confirmation of the significance, probability of the regression model using both D_{SSN} and P_{solid} was obtained for each case. The probability was then used as an independent variable for ROC analysis using binary pathological diagnosis as a dependent variable. Furthermore, the area under the curve (AUC) was calculated and compared between the semi-automatic and manual measurements [22]. P_{solid} was used instead of D_{solid} as there was a linear correlation between D_{solid} and D_{SSN} . The pathological diagnosis was divided into two groups (AAH, AIS and MIA vs. IPA) based on the distinct prognosis between them [2].

Inter- and intrareader reproducibility (agreement) of diagnosis, which was determined using a probability cut-off of 0.5 from logistic regression analysis, was investigated for each measurement method using Cohen's κ statistics. The resulting Cohen's κ values were then compared between the semi-automatic and manual measurements using Z statistics, with the null hypothesis $H_0: \kappa_1 - \kappa_2 = 0$ and the alternative hypothesis $H_a: \kappa_1 - \kappa_2 \neq 0$ [23, 24].

Inter- and intrareader reproducibility of SSN classification was also analysed using Cohen's κ statistics and was compared between the two measurement methods using Z statistics. Nodule classification was based on the Fleischner Society's recommendations for SSNs as follows: (1) pure ground-glass, (2) part-solid with a solid component 5 mm or larger or (3) less than 5 mm [2]. In addition, sub-classification of part-solid nodules was performed with D_{solid} of each reader.

The first measurement data of reader 2 (reader 2–1) was used for the calculation of inter-reader measurement variability and reproducibility of diagnosis and classification. All statistical analyses were performed using two commercial software programs (MedCalc version 12.3.0, MedCalc Software, Mariakerke, Belgium; and SPSS 19.0, IBM SPSS Statistics, Armonk, NY, USA). A p-value < 0.05 was considered to indicate statistical significance.

Results

Semi-automatic and manual diameter measurement

For reader 1, mean D_{SSN} was 14.6 ± 6.0 mm (mean \pm standard deviation) using semi-automatic measurement and 14.4 ± 6.9 mm using manual measurement ($p=0.405$). Mean D_{solid} was 6.3 ± 4.9 mm using semi-automatic measurement and 5.4 ± 4.5 mm using manual measurement ($p<0.001$). For reader 2–1, mean D_{SSN} was 14.5 ± 5.8 mm using semi-automatic measurement and 14.6 ± 6.9 mm using manual measurement ($p=0.821$). Mean D_{solid} was 6.3 ± 4.9 mm using semi-automatic measurement and 6.5 ± 5.0 mm using manual

measurement ($p=0.147$). For reader 2–2, mean D_{SSN} was 14.5 ± 5.8 mm using semi-automatic measurement and 14.1 ± 6.5 mm using manual measurement ($p=0.069$). Mean D_{solid} was 6.5 ± 5.0 mm using semi-automatic measurement and 5.9 ± 4.5 mm using manual measurement ($p<0.001$). Detailed data are provided in Table 1.

Inter- and intrareader measurement variability

For semi-automatic measurement, inter-reader variability was -1.9 mm (95 % CI: -2.3 to -1.6)– 2.1 mm (95 % CI: 1.7 – 2.4) for D_{SSN} and -2.1 mm (95 % CI: -2.5 to -1.8)– 2.1 mm (95 % CI: 1.7 – 2.5) for D_{solid} . Mean R_{diff} was 4.11 ± 8.21 % for D_{SSN} and 9.03 ± 34.02 % for D_{solid} . Intrareader measurement variability was -1.5 mm (95 % CI: -1.8 to -1.3)– 1.5 mm (95 % CI: 1.3 – 1.8) for D_{SSN} and -2.2 mm (95 % CI: -2.6 to -1.9)– 1.7 mm (95 % CI: 1.4 – 2.0) for D_{solid} . Mean R_{diff} was 2.31 ± 4.88 % for D_{SSN} and 8.89 ± 34.21 % for D_{solid} .

For manual measurement, inter-reader variability was -2.8 mm (95 % CI: -3.3 to -2.4)– 2.4 mm (95 % CI: 2.0 – 2.9) for D_{SSN} and -5.1 mm (95 % CI: -5.7 to -4.4)– 2.8 mm (95 % CI: 2.1 – 3.5) for D_{solid} . Mean R_{diff} was 7.56 ± 7.14 % for D_{SSN} and 31.27 ± 47.91 % for D_{solid} . Intrareader measurement variability was -2.1 mm (95 % CI: -2.6 to -1.7)– 3.1 mm (95 % CI: 2.7 – 3.6) for D_{SSN} and -2.1 mm (95 % CI: -2.5 to -1.6)– 3.3 mm (95 % CI: 2.8 – 3.8) for D_{solid} . Mean R_{diff} was 7.00 ± 6.57 % for D_{SSN} and 17.39 ± 34.33 % for D_{solid} . Inter- and intrareader variability of semi-automatic measurement was significantly lower than those of manual measurement for both D_{SSN} and D_{solid} (all, $p<0.001$).

Diagnostic performance

Input variables of D_{SSN} and P_{solid} were statistically significant in the logistic regression models for both semi-automatic and manual measurements of both readers ($p<0.05$; Table 2). For reader 1, the AUC was 0.907 (95 % CI: 0.834–0.956) for semi-automatic measurement and 0.833 (95 % CI: 0.747–0.900) for manual measurement ($p=0.001$). For reader 2–1, AUC was 0.905 (95 % CI: 0.831–0.954) for semi-automatic measurement and 0.862 (95 % CI: 0.780–0.922) for manual measurement ($p=0.019$). For reader 2–2, AUC was 0.923 (95 % CI: 0.853–0.967) for semi-automatic measurement and 0.864 (95 % CI: 0.782–0.924) for manual measurement ($p=0.001$). Detailed results of logistic regression analysis and ROC analysis are described in Tables 2 and 3, respectively. D_{SSN} and D_{solid} according to the pathological diagnosis, and comparisons between them can be found in Table 4.

Reproducibility of diagnosis

Inter-reader reproducibility (κ) of the diagnosis was 0.924 (95 % CI: 0.840–1.000) for semi-automatic measurement and

Table 1 Descriptive statistics for the diameter measurement

		D_{SSN} (mm)	D_{solid} (mm)	P_{solid} (%)
Reader 1	Semi-automatic measurement	14.6±6.0 (2.6, 31.1)	6.3±4.9 (0, 23.5)	38.4±23.1 (0, 78.3)
	Manual measurement	14.4±6.9 (3.7, 32.7)	5.4±4.5 (0, 22.4)	34.7±22.5 (0, 78.9)
Reader 2–1	Semi-automatic measurement	14.5±5.8 (3.8, 30.7)	6.3±4.9 (0, 24.7)	39.0±23.3 (0, 86.4)
	Manual measurement	14.6±6.9 (3.7, 34.9)	6.5±5.0 (0, 24.1)	41.8±24.1 (0, 93.8)
Reader 2–2	Semi-automatic measurement	14.5±5.8 (3.9, 30.7)	6.5±5.0 (0, 25.2)	40.8±22.8 (0, 88.1)
	Manual measurement	14.1±6.5 (4.3, 34.8)	5.9±4.5 (0, 22.9)	39.4±23.5 (0, 94.5)

Data are mean ± standard deviation. Data in parentheses are range

D_{SSN} diameter of subsolid nodule (effective diameter for semi-automatic measurement; average diameter for manual measurement), D_{solid} diameter of solid portion (effective diameter for semi-automatic measurement; average diameter for manual measurement), P_{solid} solid proportion in percentage

0.690 (95 % CI: 0.527–0.853) for manual measurement ($p=0.012$). Percentage inter-reader agreement was 97.1 % (99/102) for semi-automatic measurement and 88.2 % (90/102) for manual measurement.

Intrareader reproducibility (κ) of the diagnosis was 0.897 (95 % CI: 0.797–0.997) for semi-automatic measurement and 0.841 (95 % CI: 0.718–0.964) for manual measurement ($p=0.490$). Percentage intrareader agreement was 96.1 %

Table 2 Results of the logistic regression analysis for differentiating preinvasive lesions and minimally invasive adenocarcinomas from invasive pulmonary adenocarcinomas

		Semi-automatic measurement		Manual measurement		
		D_{SSN}	P_{Solid}	D_{SSN}	P_{Solid}	
Reader 1	Odds ratio	1.308	1.059	1.218	1.035	
	95 % CI of OR	1.118–1.530	1.027–1.092	1.086–1.367	1.009–1.061	
	p-value	0.001	<0.001	0.001	0.008	
	Model fitting					
	$Model \chi^2$	56.44		35.09		
	$p\text{-value}^a$	<0.001		<0.001		
	$Hosmer\text{-Lemeshow test}^b$	0.670		0.130		
	R^2	0.597		0.409		
Reader 2–1	Odds ratio	1.273	1.065	1.201	1.054	
	95 % CI of OR	1.094–1.482	1.033–1.099	1.072–1.347	1.025–1.084	
	p-value	0.002	<0.001	0.002	<0.001	
	Model fitting					
	$Model \chi^2$	55.93		43.96		
	$p\text{-value}^a$	<0.001		<0.001		
	$Hosmer\text{-Lemeshow test}^b$	0.518		0.553		
	R^2	0.593		0.492		
Reader 2–2	Odds ratio	1.347	1.090	1.252	1.049	
	95 % CI of OR	1.120–1.620	1.046–1.135	1.095–1.431	1.021–1.079	
	p-value	0.002	<0.001	0.001	0.001	
	Model fitting					
	$Model \chi^2$	66.41		44.13		
	$p\text{-value}^a$	<0.001		<0.001		
	$Hosmer\text{-Lemeshow test}^b$	0.429		0.596		
	R^2	0.672		0.493		

^a p-values are from the model χ^2 test

^b Data are p-values of the Hosmer-Lemeshow test.

CI confidence interval, D_{SSN} diameter of subsolid nodule (effective diameter for semi-automatic measurement; average diameter for manual measurement), OR odds ratio, P_{solid} solid proportion in percentage

Table 3 Diagnostic performance based on semi-automatic and manual measurements

		AUC ^a	Sensitivity (%)	Specificity (%)	p-value
Reader 1	Semi-automatic measurement	0.907 (0.834–0.956)	84.3	84.4	0.001
	Manual measurement	0.833 (0.747–0.900)	85.7	71.9	
Reader 2–1	Semi-automatic measurement	0.905 (0.831–0.954)	84.3	87.5	0.019
	Manual measurement	0.862 (0.780–0.922)	82.9	75.0	
Reader 2–2	Semi-automatic measurement	0.923 (0.853–0.967)	88.6	84.4	0.001
	Manual measurement	0.864 (0.782–0.924)	78.6	81.2	

^a Data in parentheses are 95 % confidence intervals. Sensitivity and specificity are obtained with an optimal cut-off based on the Youden's index

^b p-values are for the comparison between receiver operating characteristic curves of semi-automatic and manual measurements [22]

AUC area under the receiver operating characteristic curve

(98/102) for semi-automatic measurement and 94.1 % (96/102) for manual measurement.

Reproducibility of nodule classification

Inter-reader reproducibility (κ) of nodule classification was 0.861 (95 % CI: 0.769–0.953) for semi-automatic measurement and 0.683 (95 % CI: 0.561–0.805) for manual measurement ($p=0.022$) (Fig. 2). Percentage inter-reader agreement was 92.2 % (94/102) for semi-automatic measurement and 80.4 % (82/102) for manual measurement.

Intra-reader reproducibility (κ) of nodule classification was 0.894 (95 % CI: 0.812–0.976) for semi-automatic measurement and 0.750 (95 % CI: 0.632–0.868) for manual measurement ($p=0.049$). Percentage intrareader agreement was 94.1 % (96/102) for semi-automatic measurement and 85.3 % (87/102) for manual measurement. Detailed data can be found in Table 5.

Radiation dosage

Mean CTDI_{vol} and DLP were 2.22±0.90 mGy (range 0.74–5.17 mGy) and 85.27±35.61 mGy·cm (range 25.04–192.37 mGy·cm), respectively. Mean effective diameter was 25.5 ±2.0 cm (range 20.3–31.3 cm) and mean SSDE was 3.14 ±1.08 mGy (range 1.12–6.82 mGy).

Discussion

Our study results demonstrated that semi-automatic measurements were superior to manual measurements in the aspect of diagnostic accuracy for SSNs. We also demonstrated that inter-reader reproducibility of the diagnosis and classification of SSNs was significantly higher using semi-automatic measurements.

In terms of diameter measurement, D_{SSN} was shown to be similar between the semi-automatic and manual measurement methods, although there were significant differences in D_{solid} between the measurement methods for reader 1 and reader 2–2. Accordingly, the differences in D_{solid} between the pathological diagnoses were greater using semi-automatic measurements than with manual measurements. In other words, D_{solid} of preinvasive lesions and MIAs were measured smaller and/or D_{solid} of IPAs were measured larger using semi-automatic measurement, which is probably attributable to the smaller variability in the range of semi-automatic measurement of solid portions. Consequently, diagnostic performance (AUC) based on the semi-automatic effective diameter was revealed to be significantly greater than that based on the manual average diameter. To our knowledge, no studies to date have elucidated the impact of the diameter measurement method on the diagnostic performance for SSNs. Past studies focused on the measurement accuracy and variability of semi-automatic measurement. However, measurement itself is different from diagnosis. Diagnosis is based on the measurement but it is a next step in clinical process. Therefore, the plausible link between reduced measurement variability and improved diagnostic accuracy should be validated as well. The key findings of our study results are: (1) the size measurement method can affect the radiological diagnosis and (2) the accuracy of diagnosis can be augmented by using the semi-automated segmentation tool. We must note that in real clinical practice, a diagnosis is made not only using diameters but also using multiple morphological features and clinical factors. Nonetheless, diameter measurement is still the most fundamental quantitative feature in the evaluation of lung nodules.

We also hypothesized prior to our study that semi-automatic measurements would help increase the inter-reader reproducibility of a diagnosis. Given the results of previous publications on the marked improvement of measurement variability through semi-automatic segmentation

Table 4 Diameter measurements and comparisons between pathological subtypes

			Semi-automatic measurement		Manual measurement	
			AAH, AIS and MIA	IPA	AAH, AIS and MIA	IPA
Reader 1	D_{SSN}	Mean±SD	9.9±4.1	16.8±5.4	9.8±4.9	16.5±6.7
		95 % CI of mean	8.4–11.3	15.5–18.1	8.1–11.6	14.9–18.1
		p-value	<0.001		<0.001	
	D_{solid}	Mean±SD	2.0±2.4	8.2±4.6	2.3±2.6	6.8±4.5
		95 % CI of mean	1.2–2.9	7.1–9.3	1.4–3.2	5.7–7.9
		p-value	<0.001		<0.001	
	P_{solid}	Mean±SD	18.0±20.0	47.7±17.9	21.6±22.4	40.7±19.9
		95 % CI of mean	10.8–25.2	43.5–52.0	13.5–29.7	36.0–45.5
		p-value	<0.001		<0.001	
Reader 2-1	D_{SSN}	Mean±SD	10.2±4.3	16.5±5.3	10.1±5.1	16.6±6.7
		95 % CI of mean	8.6–11.7	15.3–17.8	8.2–11.9	15.0–18.2
		p-value	<0.001		<0.001	
	D_{solid}	Mean±SD	2.0±2.4	8.3±4.5	2.6±2.8	8.3±4.7
		95 % CI of mean	1.1–2.8	7.2–9.3	1.6–3.6	7.2–9.5
		p-value	<0.001		<0.001	
	P_{solid}	Mean±SD	18.0±20.2	48.6±17.8	23.7±24.9	50.0±18.8
		95 % CI of mean	10.7–25.3	44.4–52.8	14.7–32.7	45.5–54.5
		p-value	<0.001		<0.001	
Reader 2-2	D_{SSN}	Mean±SD	10.0±3.9	16.6±5.3	9.7±4.3	16.1±6.4
		95 % CI of mean	8.6–11.4	15.3–17.9	8.1–11.2	14.6–17.6
		p-value	<0.001		<0.001	
	D_{solid}	Mean±SD	2.1±2.5	8.6±4.4	2.4±2.8	7.5±4.2
		95 % CI of mean	1.2–2.9	7.5–9.7	1.4–3.4	6.5–8.5
		p-value	<0.001		<0.001	
	P_{solid}	Mean±SD	18.7±20.2	50.9±15.8	22.7±24.8	47.1±18.6
		95 % CI of mean	11.4–26.0	47.2–54.7	13.8–31.7	42.6–51.5
		p-value	<0.001		<0.001	

The unit of measurement is millimeter for D_{SSN} and D_{solid} ; and percentage (%) for P_{solid} . p-values are for comparisons between pathological subtypes (AAH, AIS and MIA vs. IPA) using either the independent t-test or Mann-Whitney U test.

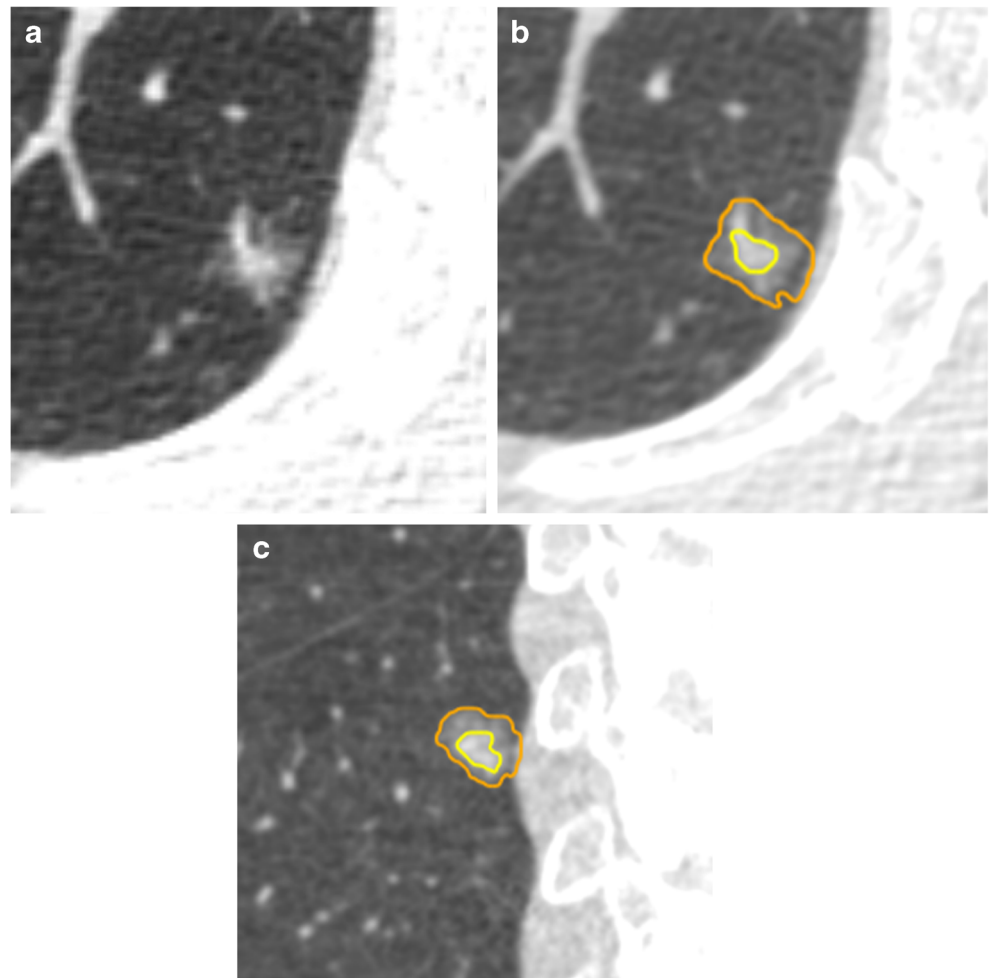
AAH atypical adenomatous hyperplasia, AIS adenocarcinoma in-situ, CI confidence interval, D_{solid} diameter of solid portion (effective diameter for semi-automatic measurement; average diameter for manual measurement), D_{SSN} diameter of subsolid nodule (effective diameter for semi-automatic measurement; average diameter for manual measurement), IPA invasive pulmonary adenocarcinoma, MIA minimally invasive adenocarcinoma, P_{solid} solid proportion in percentage, SD standard deviation

[25–27], it was deemed reasonable that semi-automatic measurement would reduce inter-reader variability of a diagnosis. As expected, the inter-reader agreement of a diagnosis was shown to be significantly higher using the semi-automatic effective diameter ($\kappa=0.924$) than with the manual average diameter ($\kappa=0.690$) in our study.

With respect to the classification of SSNs, recent studies reported substantial inter-reader variability [28–31]. Indeed, according to Riel et al. [29] who categorized nodules according to Fleischner Society guidelines, inter- and intrareader agreement was shown to be only moderate ($\kappa=0.51$ and 0.57) [29]. Penn et al. [30] and Yoo et al.

[31] also reported a similar degree of inter-reader agreement ($\kappa=0.56$ and 0.57 , respectively). Riel et al. [29] pointed out that the majority of disagreements they encountered were related to either the presence of a solid component in part-solid nodules or the size of the solid component relative to the 5-mm threshold. In our study, we demonstrated that semi-automatic measurement can help reduce inter-reader disagreement stemming from the latter factor. However, as the differentiation between pure ground-glass and part-solid nodules was performed visually in both measurement methods, we were not able to address the issue of determining the presence of a solid

Fig. 2 Measurement of the internal solid portion of an invasive pulmonary adenocarcinoma appearing as a subsolid nodule in a 77-year-old male. **(a)** Manual average diameter of the internal solid portion was measured as 4.7 mm and 7.6 mm by readers 1 and 2, respectively, on an axial plane of a non-enhanced CT scan. **(b)** Semi-automatic effective diameter of the internal solid portion was 6.6 mm for both readers. The measurement as well as nodule classification was consistent between the two readers. **(c)** Semi-automatic segmentation boundary was overlaid on the coronal CT image



portion in this study. Nevertheless, for the sub-classification of part-solid nodules according to D_{solid} using a 5-mm threshold, semi-automatic segmentation was able to significantly improve inter-reader reproducibility. Sub-classification of part-solid nodules according to D_{solid} has clinical implications considering that the management recommendations for part-solid nodules suggested by Lung CT Screening Reporting and Data System

(Lung-RADS) or the Fleischner Society, i.e. whether to suggest surveillance CT or invasive procedures (biopsy or surgical resection), are based on the solid component size (6-mm and 5-mm cut-offs, respectively) [2, 32].

It must be re-emphasized that the software package we used in our study was not fully automatic and thus needed human input to determine the nodule type and segmentation seeding point. Nevertheless, the outcome was more accurate

Table 5 Subsolid nodule classification of the two readers

		Reader 1	Reader 2–1	Reader 2–2
Semi-automatic measurement	Pure ground-glass	21 (20.6)	19 (18.6)	16 (15.7)
	Part-solid with a solid portion <5 mm	20 (19.6)	23 (22.5)	23 (22.5)
	Part-solid with a solid portion ≥5 mm	61 (59.8)	60 (58.8)	63 (61.8)
Manual measurement	Pure ground-glass	19 (18.6)	15 (14.7)	18 (17.6)
	Part-solid with a solid portion <5 mm	39 (38.2)	30 (29.4)	29 (28.4)
	Part-solid with a solid portion ≥5 mm	44 (43.1)	57 (55.9)	55 (53.9)

Data are numbers of nodules. Data in parentheses are percentages (%)

and reproducible than manual human measurement alone. In the era of lung cancer CT screening, the frequency of detected SSNs will undoubtedly increase. As pure ground-glass nodules and part-solid nodules are identified in approximately 4 % and 5 % of these baseline CT screenings [33], the need for standardization of practice for patients with SSNs will also become greater. Management guidelines such as Lung-RADS [32] or the Fleischner Society recommendations [2] contribute to relieving these concerns, and we believe that semi-automatic measurement will play an additional role in further promoting the reproducibility of nodule characterization and clinical standardization.

There are several limitations to this study. First, we only included patients who underwent surgical resections as we used their pathological diagnosis as the reference standard, and this inclusion criterion may have induced selection bias. The proportion of pure ground-glass nodules from that of all SSNs was 20.6 % (reader 1) in our study, which is lower than the expected outcome of the screening population [33]. As the variability of classification was mainly caused by the solid component measurement, the proportion of pure ground-glass and part-solid nodules might have affected the analysis results. Second, our study results may be software package-specific. As the performance of any semi-automatic measurement software depends on the segmentation algorithm used, diagnostic performance and classification reproducibility may also be dependent on the semi-automatic segmentation tool used. The segmentation success rate was 93.6 % (102/109) in our study and the optimal segmentation was not achievable in seven cases.

In conclusion, semi-automatic effective diameter measurement offers an opportunity to improve the diagnostic accuracy and reproducibility as well as classification reproducibility of SSNs, helping to facilitate a more accurate and standardized clinical practice for patients with SSNs.

Funding This study has received funding by the SNUH Research Fund, Seoul National University Hospital, Seoul, Korea (grant number: 05-2016-0050).

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Chang Min Park.

Conflict of interest The authors of this manuscript declare no relationships with any companies whose products or services may be related to the subject matter of the article.

Statistics and biometry No complex statistical methods were necessary for this paper.

Informed consent Written informed consent was waived by the Institutional Review Board.

Ethical approval Institutional Review Board approval was obtained.

Study subjects or cohorts overlap Part of the study population (36/89) had participated in a previous published study [13].

Methodology

- retrospective
- diagnostic study
- performed at one institution

References

1. Austin JH, Garg K, Aberle D et al (2013) Radiologic implications of the 2011 classification of adenocarcinoma of the lung. *Radiology* 266:62–71
2. Naidich DP, Bankier AA, MacMahon H et al (2013) Recommendations for the management of subsolid pulmonary nodules detected at CT: a statement from the Fleischner Society. *Radiology* 266:304–317
3. Lee SM, Park CM, Goo JM, Lee HJ, Wi JY, Kang CH (2013) Invasive pulmonary adenocarcinomas versus preinvasive lesions appearing as ground-glass nodules: differentiation by using CT features. *Radiology* 268:265–273
4. Kim H, Park CM, Koh JM, Lee SM, Goo JM (2014) Pulmonary subsolid nodules: what radiologists need to know about the imaging features and management strategy. *Diagn Interv Radiol* 20:47–57
5. Chae HD, Park CM, Park SJ, Lee SM, Kim KG, Goo JM (2014) Computerized texture analysis of persistent part-solid ground-glass nodules: differentiation of preinvasive lesions from invasive pulmonary adenocarcinomas. *Radiology* 273:285–293
6. Song YS, Park CM, Park SJ, Lee SM, Jeon YK, Goo JM (2014) Volume and mass doubling times of persistent pulmonary subsolid nodules detected in patients without known malignancy. *Radiology* 273:276–284
7. Ko JP, Suh J, Ibidapo O et al (2016) Lung adenocarcinoma: correlation of quantitative CT findings with pathologic findings. *Radiology* 280:931–939
8. Lee HJ, Goo JM, Lee CH et al (2009) Predictive CT findings of malignancy in ground-glass nodules on thin-section chest CT: the effects on radiologist performance. *Eur Radiol* 19:552–560
9. Takahashi S, Tanaka N, Okimoto T et al (2012) Long term follow-up for small pure ground-glass nodules: implications of determining an optimum follow-up period and high-resolution CT findings to predict the growth of nodules. *Jpn J Radiol* 30:206–217
10. Oda S, Awai K, Liu D et al (2008) Ground-glass opacities on thin-section helical CT: differentiation between bronchioloalveolar carcinoma and atypical adenomatous hyperplasia. *AJR Am J Roentgenol* 190:1363–1368
11. Gavrielides MA, Kinnard LM, Myers KJ, Petrick N (2009) Noncalcified lung nodules: volumetric assessment with thoracic CT. *Radiology* 251:26–37
12. Goo JM (2011) A computer-aided diagnosis for evaluating lung nodules on chest CT: the current status and perspective. *Korean J Radiol* 12:145–155
13. Cohen JG, Goo JM, Yoo RE et al (2016) The effect of late-phase contrast enhancement on semiautomatic software measurements of CT attenuation and volume of part-solid nodules in lung adenocarcinomas. *Eur Radiol* 85:1174–1180
14. American Association of Physicists in Medicine (2011) Size-specific dose estimates (SSDE) in pediatric and adult body CT examinations. Task Group. American Association of Physicists in Medicine, College Park, p 204
15. Cohen JG, Goo JM, Yoo RE et al (2016) Software performance in segmenting ground-glass and solid components of subsolid nodules in pulmonary adenocarcinomas. *Eur Radiol* 26:4465–4474

16. Scholten ET, Jacobs C, van Ginneken B et al (2015) Detection and quantification of the solid component in pulmonary subsolid nodules by semiautomatic segmentation. *Eur Radiol* 25:488–496
17. Kuhnigk J-M, Dicken V, Bornemann L et al (2006) Morphological segmentation and partial volume analysis for volumetry of solid pulmonary lesions in thoracic CT scans. *IEEE Trans Med Imaging* 25:417–434
18. de Hoop B, Gietema H, van Ginneken B, Zanen P, Groenewegen G, Prokop M (2009) A comparison of six software packages for evaluation of solid lung nodules using semi-automated volumetry: what is the minimum increase in size to detect growth in repeated CT examinations. *Eur Radiol* 19:800–808
19. Lee KH, Goo JM, Park SJ et al (2014) Correlation between the size of the solid component on thin-section CT and the invasive component on pathology in small lung adenocarcinomas manifesting as ground-glass nodules. *J Thorac Oncol* 9:74–82
20. Travis WD, Brambilla E, Noguchi M et al (2011) International association for the study of lung cancer/American thoracic society/European respiratory society international multidisciplinary classification of lung adenocarcinoma. *J Thorac Oncol* 6:244–285
21. Bland JM, Altman DG (1999) Measuring agreement in method comparison studies. *Stat Methods Med Res* 8:135–160
22. DeLong ER, DeLong DM, Clarke-Pearson DL (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 44:837–845
23. Congalton RG, Green K (2008) Assessing the accuracy of remotely sensed data: Principles and practices. CRC press, Boca Raton
24. de Leeuw J, Jia H, Yang L, Liu X, Schmidt K, Skidmore A (2006) Comparing accuracy assessments to infer superiority of image classification methods. *Int J Remote Sens* 27:223–232
25. Scholten ET, de Hoop B, Jacobs C et al (2013) Semi-automatic quantification of subsolid pulmonary nodules: comparison with manual measurements. *PLoS One* 8:e80249
26. Xie X, Zhao Y, Snijder RA et al (2013) Sensitivity and accuracy of volumetry of pulmonary nodules on low-dose 16- and 64-row multi-detector CT: an anthropomorphic phantom study. *Eur Radiol* 23:139–147
27. Zhao B, James LP, Moskowitz CS et al (2009) Evaluating variability in tumor measurements from same-day repeat CT scans of patients with non-small cell lung cancer. *Radiology* 252:263–272
28. Ridge CA, Yildirim A, Boiselle PM et al (2016) Differentiating between subsolid and solid pulmonary nodules at CT: inter- and intraobserver agreement between experienced thoracic radiologists. *Radiology* 278:888–896
29. van Riel SJ, Sanchez CI, Bankier AA et al (2015) Observer variability for classification of pulmonary nodules on low-dose CT images and its effect on nodule management. *Radiology* 277:863–871
30. Penn A, Ma M, Chou BB, Tseng JR, Phan P (2015) Inter-reader variability when applying the 2013 Fleischner guidelines for potential solitary subsolid lung nodules. *Acta Radiol* 56:1180–1186
31. Yoo RE, Goo JM, Hwang EJ et al (2017) Retrospective assessment of interobserver agreement and accuracy in classifications and measurements in subsolid nodules with solid components less than 8mm: which window setting is better? *Eur Radiol* 27:1369–1376
32. American College of Radiology (2014) Lung CT Screening Reporting and Data System (Lung-RADS). American College of Radiology. Available via www.acr.org/Quality-Safety/Resources/LungRADS. Accessed 15 Oct 2016
33. Yankelevitz DF, Yip R, Smith JP et al (2015) CT screening for lung cancer: nonsolid nodules in baseline and annual repeat rounds. *Radiology* 277:555–564