Detection of Subsolid Nodules in Lung Cancer Screening
Complementary Sensitivity of Visual Reading and Computer-Aided Diagnosis

Mario Silva, MD, PhD,† Cornelia M. Schaefer-Prokop, MD, PhD,‡§ Colin Jacobs, PhD,‡ Giovanni Capretti, MD,* Francesco Ciompi, PhD,‡ Bram van Ginneken, PhD,‡ Ugo Pastorino, MD,† and Nicola Sverzellati, MD, PhD**

Objectives: The aim of this study was to compare computer-aided diagnosis (CAD) and visual reading for the detection of subsolid nodules (SSNs) in volumetrl measuremic low-dose computed tomography (LDCT) for lung cancer screening.

Materials and Methods: Prospective visual detection (VD) and manuaent of SSN were performed in the 2303 baseline volumetric LDCTs of the Multicenter Italian Lung Detection trial. Baseline and 2- and 4-year LDCTs underwent retrospective CAD analysis, subsequently reviewed by 2 experienced thoracic radiologists. The reference standard was defined by the cumulative number of SSNs detected by any reading method between VD and CAD. The number of false-positive CAD marks per scan (FP^{SSN}/scan) was calculated. The positive predictive value of CAD was quantified per nodule (PPV^{SSN}) and per screenee ($PPV^{Screenee}$). The sensitivity and negative predictive value were compared between CAD and VD. The longitudinal 3-time-point sensitivity of CAD was calculated in the subgroup of persistent SSNs seen by VD (ratio between the prevalent SSNs detected by CAD through 3 time points and the total number of persistent prevalent SSNs detected by VD) to test the sensitivity of iterated CAD analysis during a screening program. Semiautomatic characteristics (diameter, volume, and mass; both for whole nodule and solid component) were compared between SSN detected CAD-only or VD-only to investigate whether either reading method could suffer from specific sensitivity weakness related to SSN features. Semiautomatic and manual diameters were compared using Spearman ρ correlation and Bland-Altman plot.

Results: Computer-aided diagnosis and VD detected a total of 194 SSNs in 6.7% (155/2,303) of screenees at baseline LDCT. The CAD showed mean FP^{SSN}/scan of 0.26 (604/2,303); PPV^{SSN} 22.5% (175/779) for any SSN, with 54.4% (37/68) for PSN and 19.4% for NSN (138/711; $P < 0.001$); $PPV^{Screenee}$ 25.6% (137/536). The sensitivity of CAD was superior to that of VD (88.4% and 34.2%, $P < 0.001$), as well as negative predictive value (99.2% and 95.5%, $P < 0.001$). The longitudinal 3-time-point sensitivity of CAD was 87.5% (42/48). There was no influence of semiautomatic characteristics on the performance of either reading method. The diameter of the solid component in PSN was larger by CAD compared with manual measurement. At baseline, CAD detected 3 of 4 SSNs, which were first overlooked by VD and subsequently evolved to lung cancer.

Conclusions: Computer-aided diagnosis and VD as concurrent reading methods showed complementary performance, with CAD having a higher sensitivity, especially for PSN, but requiring visual confirmation to reduce false-positive calls. Computer-aided diagnosis and VD should be jointly used for LDCT reading to reduce false-negatives of either lone method. The semiautomatic measurement of solid core showed systematic shift toward a larger diameter, potentially resulting in an up-shift within Lung CT Screening Reporting and Data System classification.

Key Words: computer-aided diagnosis, lung cancer screening, low-dose computed tomography, subsolid nodule, nonsolid nodule, ground-glass nodule, part-solid nodule, detection, early diagnosis, risk stratification

(*Invest Radiol* 2018;00: 00–00)

The prognostic relevance of pulmonary subsolid nodules (SSNs) has been unveiled and supported by lung cancer screening trials throughout the last 20 years. Subsolid nodules are less frequent than solid nodules but have a higher rate of malignancy (19%–79%).[1,2] Most importantly, they represent early lung cancer stages and, mostly, slow-growing tumors.[3–5] In most lung cancer screening trials by low-dose computed tomography (LDCT), a substantial proportion of cancer-harboring SSNs were overlooked or misinterpreted (eg, bizarre nonnodular shape and nodules too small to classify) by visual reading at early rounds, albeit retrospectively visible.[6,7]

Within the work-up of lung cancer screening, computer-aided diagnosis (CAD) systems are increasingly used to analyze volumetric LDCT for detection of nodules, thereby increasing sensitivity and reducing reading time.[8–11] Computer-aided diagnosis was first developed and commercially available only for solid nodules, because the importance of solid nodules for early lung cancer detection was well known for decades. Furthermore, the development of software was much more straightforward for solid nodules because the high attenuation difference between solid nodule and surrounding lung parenchyma allowed for secure automatic border detection.[9] Conversely, detection of SSN is more challenging both by visual reading and CAD analysis.

The relatively low attenuation difference and the less well defined lesion margins are challenging for automatic detection by density-based analytics.[12] The introductory phantom-based investigations of CAD for detection of SSN[13] were followed by validation on selected subjects with visually detected SSN.[14,15] This approach, however, did not allow for evaluating whether CAD was able to detect SSN overlooked by visual reading. To our knowledge, the performance of CAD as concurrent reader for SSN has not been tested yet. These data are needed in the perspective of large workloads deriving from population-based lung cancer screening.

The objective of this study was, first, to compare the performance between CAD and visual reading for the detection of SSN, using the complete LDCT dataset of the Multicenter Italian Lung Detection (MILD) lung cancer screening trial. Second, we aimed to quantify the difference between semiautomatic and manual measurements for characterization of SSN.

Received for publication November 29, 2017; and accepted for publication, after revision, February 12, 2018.

From the *Section of Radiology, Unit of Surgical Sciences, Department of Medicine and Surgery (DiMeC), University of Parma, Parma, Italy; †Department of Thoracic Surgery, IRCCS Istituto Nazionale Tumori, Milan, Italy; ‡Department of Radiology and Nuclear Medicine, Radboud University Medical Center, Nijmegen, the Netherlands; §Department of Radiology, Meander Medical Center, Amersfoort, the Netherlands.

Conflicts of interest and source of funding: The Multicenter Italian Lung Detection project was supported by grants from the Italian Ministry of Health (RF 2004), the Italian Association for Cancer Research (AIRC 2004 IG 1227 and AIRC 5xmille IG 12162), and the Fondazione Cariplo (2004-1560). CJ received research grants from MeVis Medical Solutions AG.

The authors report no conflicts of interest.

Correspondence to: Mario Silva, MD, PhD, Section of Radiology, Unit of Surgical Sciences, Department of Medicine and Surgery (DiMeC), University of Parma, Pad. Barbieri, via Gramsci 14, 43126 Parma, Italy. E-mail: mario.silva@unipr.it.

ISSN: 0020-9996/18/0000–0000
DOI: 10.1097/RLI.0000000000000464

MATERIALS AND METHODS

Population

This study was performed in the context of the MILD trial (www.clinicaltrials.gov Identifier: NCT02837809), which was approved by the local institutional review board. Written informed consent was obtained from all screenees at study entry and included retrospective analysis of the prospectively acquired data. The present study comprised all 2303 screenees randomized to the LDCT arm between September 2005 and January 2011 (mean age, 58.1 ± 5.9 years; range, 49–78 years; 1570 men: mean age, 58.4 ± 5.9 years, range, 49–78 years; 733 women: mean age, 57.3 ± 5.9 years, range, 49–76 years).[16] All volumetric LDCTs were performed using a 16-detector-row computed tomography (CT) scanner (Siemens, Forchheim, Germany) during 1 deep inspiratory breath-hold, with the following parameters: tube voltage 120 kV, tube current 30 mAs, collimation 0.75 mm, pitch 1.5, rotation time 0.5 seconds, reconstructed slice thickness 1 mm, reconstruction increment 1 mm, medium-sharp kernel (B50f), and lung window (window width 1600 HU, window level −600 HU).[17]

Detection of SSN

Within the prospective execution of the screening trial, all baseline LDCTs had been visually evaluated by 1 of the 7 screening radiologists (4–20 years of experience in reading thoracic CT) for detection of SSN, including nonsolid nodules (NSNs) and part-solid nodules (PSNs). No CAD was used for detection of SSN. These data will be referred to as VD (visual detection).

For the purpose of this study, an advanced lung screening workstation with integrated CAD (CIRRUS Lung Screening, Nijmegen, the Netherlands) was used to retrospectively read the LDCT scans of the baseline, and 2 incidence rounds (2 and 4 years after baseline). The 2 incidence rounds were analyzed to investigate the ability of repeated CAD analysis for postponed detection of persistent SSN missed at baseline. The used workstation is a prototype version of Veolity (MeVis Medical Solutions AG, Bremen, Germany), which is a Food and Drug Administration and European Community approved reading workstation for lung cancer screening scans. Beyond Veolity tools, the present prototype of CAD software includes an algorithm for detection of SSN.[18] All CAD marks were jointly reviewed by 2 experienced thoracic radiologists (8 and 11 years of experience in screening LDCT), who selected CAD marks reflecting SSNs (TP^{SSN}) and discarded false-positive CAD marks (FP^{SSN}). The 2 radiologists were blinded to the VD results. These data will be referred to as CAD.

For data analysis, SSNs were categorized into (*a*) detected only by CAD (CAD-only), (*b*) detected only by VD (VD-only), and (*c*) detected by both CAD and VD.

Manual Measurement of Diameter

Manual measurements were obtained for each nodule detected by VD, including whole nodule diameter (D^{Manual}) and solid component ($_{solid}D^{Manual}$, if present), according to the Fleischner Society recommendations for measuring pulmonary nodules at CT.[19] The D^{Manual} was calculated as the mean of the maximum nodule diameter and its orthogonal diameter by electronic calipers. In PSN, the $_{solid}D^{Manual}$ was calculated as the maximum diameter of the solid component by electronic caliper.

Semiautomatic Quantitative Analysis

The CAD software performed semiautomatic segmentation of the SSN using predefined threshold density values, both for the whole nodule outer borders of the non-solid component (threshold >−750 HU) and for the solid component (threshold >−300 HU).[20]

The initial segmentation was based on region growing with density thresholds optimized to LDCT features of SSN, followed by a dedicated sequence of morphological operations to remove adjacent structures such as vessels and the pleural wall from the nodule. The reviewing experienced thoracic radiologists could refine the SSN segmentation by editing 2 parameters: (*a*) *density threshold*, which adjusts the density thresholds of the algorithm, and (*b*) *roundness versus irregularity*, which determines the aggressiveness of the morphological operations.[20] This procedure was applied to the whole SSN segmentation and to the segmentation of the solid core (if present).

The quantitative measurement was performed for all SSNs, including VD-only SSNs. The following semiautomatic parameters were calculated: (*a*) effective diameter defined as the diameter of a geometrical sphere with the segmented volume (D^{CAD} in mm), (*b*) whole volume (V in mm^3), and (*c*) whole mass (M in mg).[21] All 3 measures were obtained separately also for the solid component, if present, namely, (*d*) effective solid diameter ($_{solid}D^{CAD}$ in mm), (*e*) solid volume ($_{solid}V$ in mm^3), and (*f*) solid mass ($_{solid}M$ in mg).

The clinical value of differences between manual and semiautomatic quantitative measurement were tested according to Lung CT Screening Reporting and Data System (Lung-RADS) scoring system.

Statistical Analysis

The size threshold for detection of SSN was set at 5 mm, according to the histologic evidence that SSN above this size shows cancerous potential.[22] The reference standard was defined by the cumulative number of SSNs detected by any reading method between VD and CAD.

Visual Review of CAD Marks: Distribution of Positives

The distribution of true-positive (TP) and false-positive (FP) CAD marks at baseline was analyzed using 3 methodological approaches, respectively, at the level of the nodule, the CT scan, and the screenee.[23]

To quantify the *nodule-wise* performance, we calculated the positive predictive value for SSN (PPV^{SSN}), defined as the ratio between TP^{SSN} and all CAD marks ($TP^{SSN} + FP^{SSN}$).

To quantify the *scan-wise* performance, we determined the mean number of FP^{SSN} per CT scan (FP^{SSN}/scan), calculated as the ratio between the total number of FP^{SSN} at baseline and the total number of baseline scans (n = 2303).

To quantify the *screenee-wise* performance, we determined the PPV per screenee ($PPV^{Screenee}$) as the ratio between screenees with at least 1 TP CAD mark and the total number of screenees with at least 1 CAD mark (TP or FP).

Detection of the Risk-Dominant SSN

In screenees showing 1 or several SSNs either by CAD or VD or both, the risk-dominant SSN was defined as the SSN with the highest rank by the Lung-RADS.[24] Subsequently, the sensitivity (SENS) and negative predictive value (NPV) of CAD and VD for the risk-dominant nodule were calculated. The McNemar χ^2 test was used to test significance of differences.

The rate of agreement between CAD and VD results was calculated by Cohen κ coefficient (poor agreement, ≤0.2; fair, 0.2-0.4; moderate, 0.4-0.6; good, 0.6-0.8; excellent, 0.8-1).[25]

To assess whether nodule type or nodule size had an impact on the performance of either reading method on baseline performance, we compared the semiautomatic parameters of CAD-only and VD-only SSNs as determined by semiautomatic measurements. Significance of difference was tested using the Mann-Whitney *U* test. Continuous variables were reported as median and 25th and 75th percentile.

In the subgroup of persistent SSNs seen by VD, the longitudinal 3-time-point sensitivity of CAD was calculated as the ratio between the SSN detected by CAD through the 3 time points (eg, through baseline, 2-year, and 4-year round) and the total number of persistent SSNs detected by VD. This figure provides a close-to-reality simulation of

CAD performance in a longitudinal lung cancer screening program with at least 3 time points.

Comparison of Semiautomatic and Manual Diameters

For SSNs detected by VD, the correlation between semiautomatic (eg, D^{CAD} and $_{solid}D^{CAD}$) and manual (eg, D^{Manual} and $_{solid}D^{Manual}$) diameters was tested by the Spearman ρ. The variability between semiautomatic and manual diameters was tested by the Bland-Altman plot. Systematic differences were reported as mean ± SD. The absolute percentage error was reported as median with 95% confidence interval (95% CI). The limits of agreement were reported as lower and upper limit with their respective 95% CI. The rate of agreement for Lung-RADS score between manual and semi-automatic measurement was calculated by the weighted κ coefficient.[25] The manual measurement was used as reference to determine the category shift by semiautomatic measurement.

Statistically significant difference was determined at $P < 0.05$. The statistical analysis was performed by MedCalc Statistical Software version 17.4 (MedCalc Software bvba, Ostend, Belgium; https://www.medcalc.org; 2017).

RESULTS

A total of 194 SSNs were detected in 155 of 2303 (6.7%) screenees at baseline, including CAD and VD findings; this amount served as research reference standard. A total of 130 had a single SSN and 25 screenees had multiple SSNs (median, 2 SSNs, range, 2–6 SSNs). Visual detection detected 72 of the 194 SSNs (37.1%), and CAD detected 175 SSNs (90.2%); the segmentation was edited in 22 of 175 (13%) SSNs.

Visual Review of CAD Marks: Distribution of Positives

Per nodule, the PPV for SSNs was 22.5% (175/779), being significantly higher for PSN (54.4%, 37/68) compared with NSN (19.4%, 138/711; $P < 0.001$).

Per scan, CAD showed a mean of 0.26 FP^{SSN}/scan (604/2303). In detail, 0.25 FP/scan was seen for NSN (573/2303) and only 0.01 FP/scan for PSN (31/2303).

Per screenee, CAD found 536 screenees with at least 1 SSN candidate, which were subsequently classified by visual review as 137 TPs and 399 FPs. The $PPV^{Screenee}$ was 25.6% (137/536).

Detection of the Risk-Dominant SSN

The risk-dominant SSN was an NSN in 78% (121/155) of screenees and a PSN in 22% (34/155) of screenees. Computer-aided diagnosis showed a significantly higher SENS (88.4%) and NPV (99.2%) than VD (SENS, 34.2%, and NPV, 95.5%). The superior SENS and NPV of CAD were true for both the NSN and, especially, the PSN. Table 1 details SSNs according to method of detection, nodule density, SENS, and NPV.

The overall agreement was only fair between CAD and VD for the detection of SSNs ($\kappa = 0.246$, 95% CI, 0.135-0.357; $P < 0.001$). Four SSNs were resected with a histological diagnosis of adenocarcinoma (range of time from baseline LDCT to histological diagnosis was 4 to 7 years); all of them were detected by CAD at baseline, whereas only 1 had been visually seen at baseline.

Nodule type—NSN or PSN—was not a determinant of detection failure by either method ($P = 0.235$). Semiautomatic quantitative parameters of whole nodule and solid core were not significantly different between CAD-only and VD-only detected SSNs, with the limitation of a small sample of VD-only PSN (n = 2; Table 2).

On baseline scan, CAD showed a sensitivity of 66% (35/53) in the subgroup of SSNs seen by VD. From among SSNs seen by VD at baseline, 48 of 53 (91%) were persistent and 5 of 53 (9%) were transient at the following LDCT rounds. On longitudinally iterated CAD analysis, 7 SSNs were detected by CAD at the 2-year (3 NSNs; Fig. 1) or 4-year round (4 NSNs; Fig. 2), whereas 6 were never detected by CAD (5 NSNs and 1 PSN; Fig. 3). This lead to a longitudinal 3-timepoint sensitivity of 87.5% (42/48) for CAD in the subgroup of persistent SSNs seen by VD. None of the 6 SSNs iteratively overlooked by CAD evolved to lung cancer.

Comparison Between Semiautomatic and Manual Diameters

In the 53 SSNs detected by VD, there was significant correlation between D^{CAD} and D^{Manual} (Spearman $\rho = 0.752$, 95% CI, 0.605–0.850;

TABLE 1. Risk-Dominant SSNs Detailed According to Method of Detection and Nodule Density

	All SSNs	NSN	PSN
		Detection, N (%)	
Total	155 (100%)	121 (100%)	34 (100%)
CAD-only	102 (65.8%)	75 (62.0%)	27 (79.4%)
CAD and VD	35 (22.6%)	30 (24.8%)	5 (14.7%)
VD-only	18 (11.6%)	16 (13.2%)	2 (5.9%)
		Sensitivity, % (TP/TP + FN)	
CAD	88.4% (137/155)	86.8% (105/121)	94.1% (32/34)
VD	34.2% (53/155)	38% (46/121)	20.6% (7/34)
P	<0.001	<0.001	<0.001
		Negative Predictive Value, % (TN/TN + FN)	
CAD	99.2% (2,148/2,166)	99.3% (2,182/2,198)	99.9% (2,269/2,271)
VD	95.5% (2,148/2,250)	96.7% (2,182/2,257)	98.8% (2,269/2,296)
P	<0.001	<0.001	<0.001

The sensitivity and negative predictive value is compared between CAD and VD for all nodules and in the subgroups of NSN and PSN. Sensitivity and negative predictive value of CAD are significantly higher compared with VD, for all SSNs and for each of NSN and PSN.

CAD indicates computer-aided diagnosis; FN, false-negative; NSN, nonsolid nodule; PSN, part-solid nodule; SSN, subsolid nodule; TN, true-negative; TP, true-positive; VD, visual detection.

TABLE 2. Semiautomatic Parameters According to Detection Method (either CAD-Only or VD-Only) of Risk-Dominant SSN

	SSNs	CAD-Only Detected SSNs	VD-Only Detected SSNs	
	155 (100%)	102/155 (65.8%)	18/155 (11.6%)	*P*
D^{CAD} (mm)	7.2 (6.1–9.7)	7.2 (6.0–9.2)	8.0 (7.0–9.0)	0.252
V (mm^3)	192.3 (120.9–477.2)	190.4 (114.1–407.0)	176.2 (119.6–623.2)	0.669
M (mg)	82.2 (49.5–208.4)	76.2 (51.4–184.0)	80.7 (39.2–189.9)	0.499
	NSNs	**CAD-Only Detected NSNs**	**VD-Only Detected NSNs**	
	121/155 (78.1%)	75/121 (62.0%)	16/121 (13.2%)	
D^{CAD} (mm)	7.1 (6.1–9.4)	6.8 (6.0–8.9)	6.4 (6.1–11.1)	0.798
V (mm^3)	189.2 (120.3–431.5)	167.9 (114.6–337.5)	137.9 (119.0–709.8)	0.739
M (mg)	76.3 (44.4–174.1)	71.6 (46.5–147.1)	51.6 (39.1–163.4)	0.515
	PSNs	**CAD-Only Detected PSNs**	**VD-Only Detected PSNs**	
	34/155 (21.9%)	27/34 (79.4%)	2/34 (5.9%)	
D^{CAD}(mm)	8.2 (6.3–11.5)	7.8 (6.1–11.2)	8.9 (ND*)	ND*
V (mm^3)	288.7 (133.7–789.8)	253.3 (118.6–737.5)	374.7 (ND*)	ND*
M (mg)	126.3 (66.7–353.1)	120.6 (58.5–324.7)	158.5 (ND*)	ND*
$_{solid}D^{CAD}$ (mm)	4.2 (2.9–6.0)	4.0 (2.8–5.7)	3.8 (ND*)	ND*
$_{solid}V$ (mm^3)	39.2 (13.2–111.5)	33.6 (11.5–97.7)	30.3 (ND*)	ND*
$_{solid}M$ (mg)	36.2 (12.1–102.4)	32.8 (10.5–92.3)	26.5 (ND*)	ND*

Values are reported as median and 25th and 75th percentile.

*The VD-only detected group was extremely small (n = 2).

CAD indicates computer-aided diagnosis; ND, not defined; NSN, nonsolid nodule; PSN, part-solid nodule; SSN, subsolid nodule; VD, visual detection.

$P < 0.001$; Fig. 4A). The comparison of D^{CAD} and D^{Manual} showed a systematic difference of −0.55±1.9 mm corresponding to an absolute percentage error of 13.3% (95% CI, 10.5%–16.8%), the lower limit of agreement was −4.4 mm (95% CI, −5.4 to −3.5), and the upper limit of agreement 3.3 mm (95% CI, 2.4–4.3). The Bland-Altman plot showed no net overgoing or underrating of nodule size ($P = 0.374$; Fig. 4B).

In the subset of 7 PSNs detected by VD, there was no significant correlation between $_{solid}D^{CAD}$ and $_{solid}D^{Manual}$ (Spearman $\rho = 0.668$, 95% CI, −0.171 to 0.945; $P = 0.101$; Fig. 5A). The comparison of $_{solid}D^{CAD}$ and $_{solid}D^{Manual}$ showed a systematic difference of −1.6 ± 1.5 mm (range corresponding to an absolute percentage error of 24% [95% CI, 5.9–47.4]), a lower limit of agreement of −4.5 mm (95% CI, −6.9 to −2.1), and an upper limit of agreement 1.2 mm (95% CI, −1.2 to 3.7). The Bland-Altman plot showed that the diameter of the solid component in PSN was larger by CAD compared with manual measurement ($P = 0.026$; Fig. 5B).

The rate of agreement between semiautomatic and manual diameters for Lung-RADS category was 94% ($\kappa = 0.7523$, 95% CI, 0.5753–0.9294; $P < 0.0001$). Lung-RADS category variation was seen in 3 of 53 (6%) SSNs, all of them with category up-shift by semiautomatic measurement (Table 3). In particular, all NSNs were the same category by semiautomatic and manual measurement, whereas rank up-shift happened by semiautomatic diameter in as many as 3 of 7 (43%) PSNs. With the limit of the small number of PSNs, the rate of agreement for Lung-RADS category for PSN was 57% (κ = na).

DISCUSSION

Our results indicate that CAD had higher sensitivity than visual reading (88.4%, 137/155 vs 34.2%, 53/155). As a consequence, CAD found more screenees with SSN compared with the visual reading (5.9%, 137/2,303 vs 2.3%, 53/2,303). In 3 screenees, adenocarcinoma developed from SSN that had been originally missed by visual reading

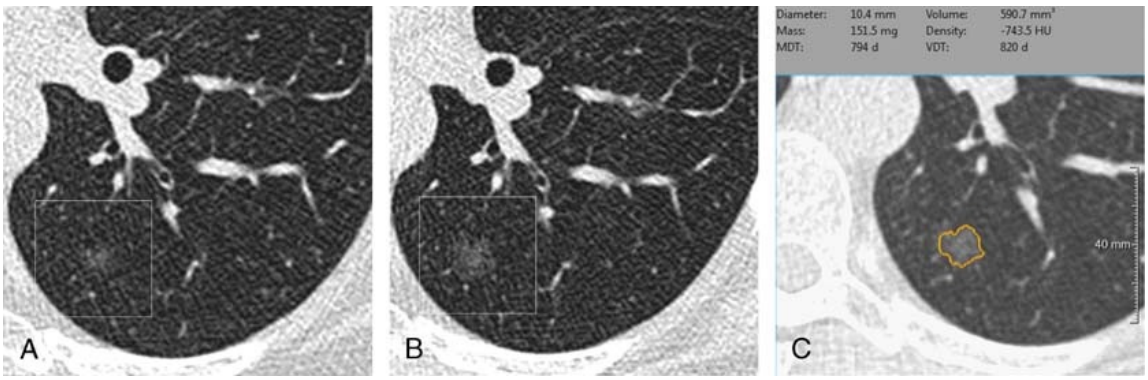

FIGURE 1. A–C, Axial noncontrast LDCT scan of persistent NSN detected by VD at baseline and by CAD at the 2-year round (54-year old woman). Baseline LDCT (A) shows the NSN at the time of CAD overlooking. The 2-year LDCT (B) shows the NSN at the time of CAD detection and its semiautomatic segmentation along with semiautomatic parameters (C—the orange contour outlines the whole segmented volume). MDT indicates mass doubling time; VDT, volume doubling time.

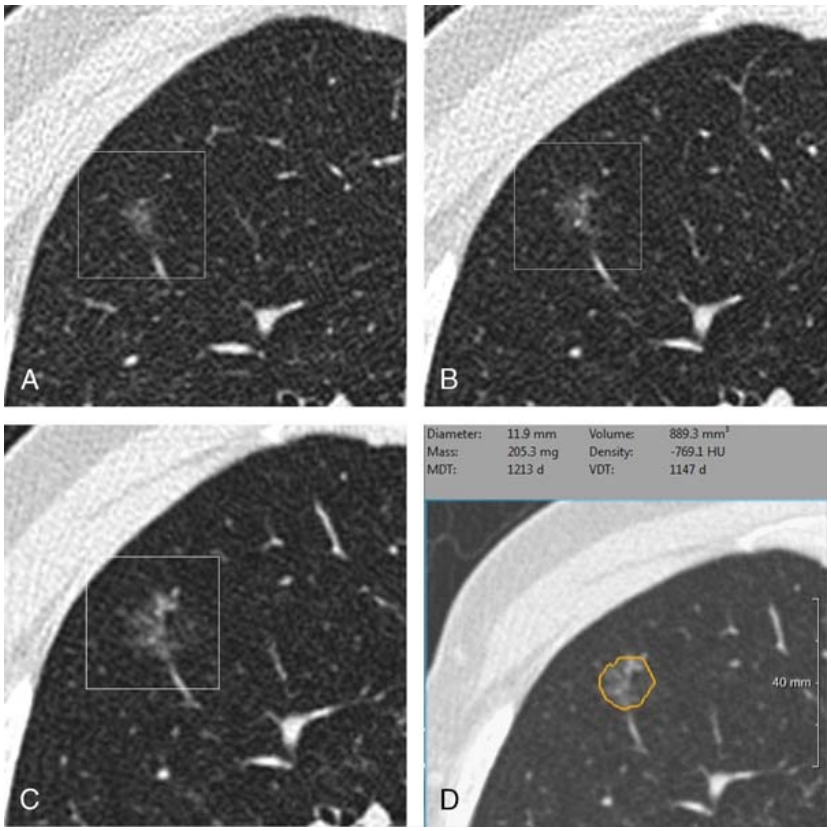

FIGURE 2. A–D, Axial noncontrast LDCT scan of persistent NSN detected by VD at baseline and by CAD at the 4-year round (55-year old man). Baseline (A) and 2-year (B) LDCTs show the NSN at the time of CAD overlooking. The 4-year LDCT (C) shows the NSN at the time of CAD detection and its semi-automatic segmentation along with semi-automatic parameters (D—the orange line outlines the whole segmented volume. MDT, mass doubling time; VDT, volume doubling time).

on the baseline scan. The higher detection rate of CAD came at the cost of 0.26 FP/scan.

On the other hand, CAD detected only 66% (35/53) of the SSNs seen by VD, indicating that visual and computer-enhanced analysis are complementary. Computer-aided diagnosis alone was found to be not sufficiently accurate to replace additional visual analysis, even when some of the SSNs missed in baseline scans were correctly detected in follow-up scans.

A number of previous studies have analyzed the performance of CAD for the detection of SSNs selected by visual reading.[14,18,26–30] Reported sensitivities ranged from 50% to 90% at the cost of 2 to 17 FP/scan, including standard dose CT and LDCT. The CAD system tested in our article had been previously evaluated by Jacobs et al,[18] who found a sensitivity of 80% at 1 FP^{SSN}/scan in a selection of screenees with visually detected SSNs from the Dutch-Belgian Lung Cancer Screening trial. Based on the free response receiver operating characteristic, the authors reported a sensitivity of 60% at 0.25 FP/scan, which is comparable with our results in the subgroup of SSNs reported by visual reading (66% sensitivity at 0.26 FP^{SSN}/scan). This balance between good sensitivity and low FP rate is better than previous reports,[14,18,26–29] and selection algorithms are continuously developed for further reduction of FP rate.[31]

Our results expand the current literature by reporting the large amount of risk-dominant SSNs that were overlooked by visual reading but detected by CAD analysis (102/155; namely, 65.8%). Previous studies showed that the sensitivity of CAD is higher for PSN compared with NSN on standard-dose CT.[14,26] We confirm such performance also on LDCT from an unselected screening population: almost 80%

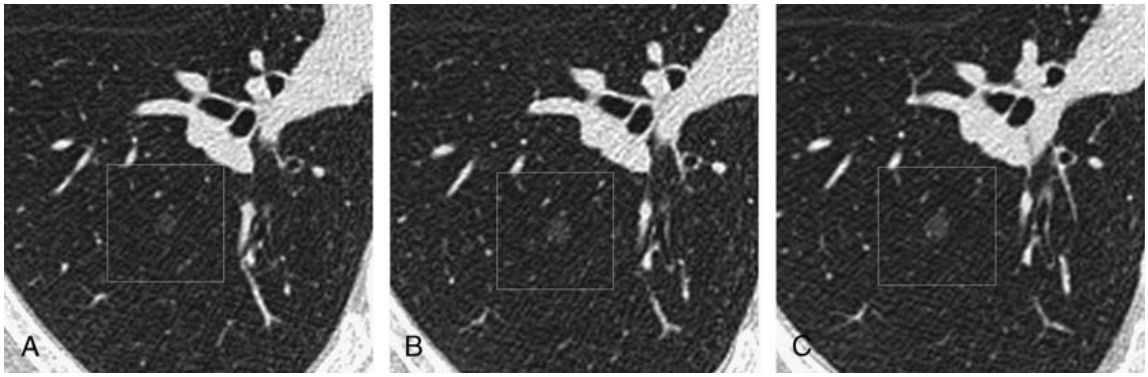

FIGURE 3. A–C, Axial noncontrast LDCT scan of persistent NSN detected only by VD at baseline and repeatedly overlooked by CAD (57-year old man). Baseline (A), 2-year (B), and 4-year (C) LDCTs show the persistent NSN, which was overlooked by CAD.

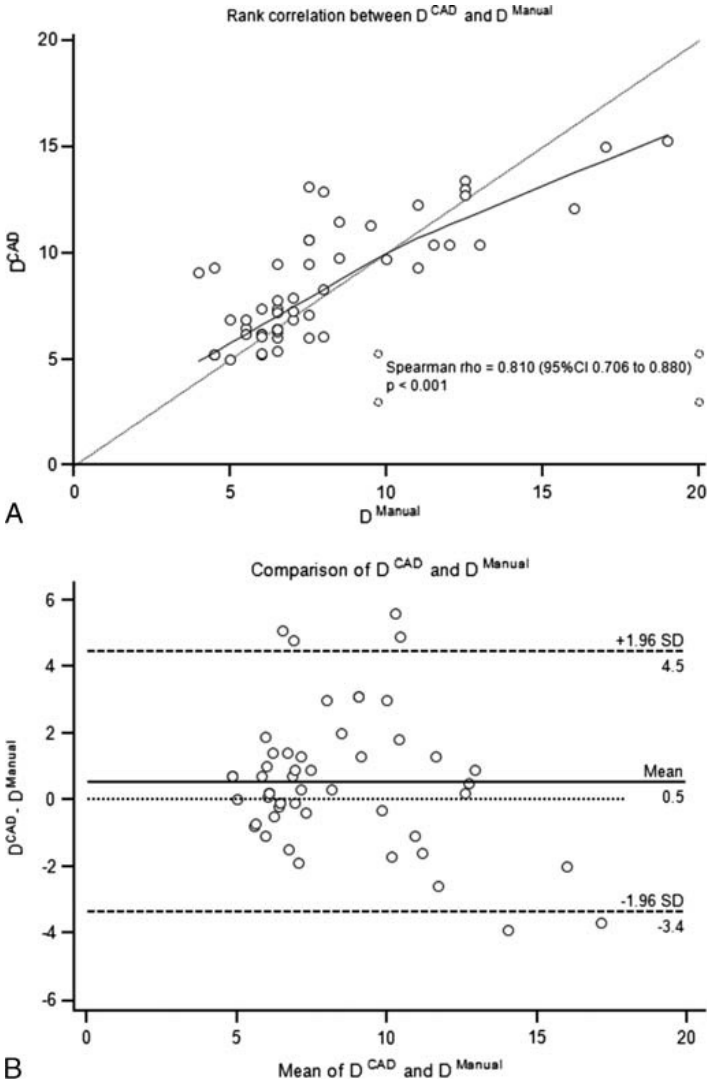

FIGURE 4. A–B, Scatter diagram of rank correlation (A) and Bland-Altman plot (B) for comparison between D^{CAD} and D^{Manual}.

PSNs were detected only by CAD. This high capability of CAD to detect PSN is remarkable in the context of lung cancer screening because PSN is much more likely than NSN to harbor foci of stromal invasion in asymptomatic adenocarcinoma[22]; hence, its detection is more clinically relevant.

The performance of a CAD system is usually quantified as the PPV, describing the proportion of TP CAD marks confirmed after visual review—either at the level of nodule or of screenee—compared with all CAD marks. In a subpopulation from the Dutch-Belgian Lung Cancer Screening trial (random selection of 400 screenees), Zhao et al[8] described a nodule-wise PPV of 8.9% for any nodule and of 16.2% for nodules 50 mm^3 or bigger. However, the specific PPV for SSN has thus far not been reported. In our study, CAD showed a PPV^{SSN} of 22.5% (175/779). Because SSN might be multiple in a single screenee, we sought to report also the screenee-wise PPV for SSN. Noteworthy, the screenee-wise approach is supported by the growing evidence that the likelihood of lung cancer is proportional to the LDCT findings, and screenees without any LDCT findings could be assigned to a longer follow-up interval, for instance, biennial rounds to save radiation dose and reduce harms, workload, and costs of screening.[16,33] We report a $PPV^{Screenee}$ of 25.6% (137/536); namely, about 3 of 4 screenees with CAD mark were defined negative after visual review. These data suggest that complementary visual review by radiologists is still mandatory for differentiating true from FP CAD marks.

In this study, multiple-time-point LDCTs were analyzed by CAD for each screenee (baseline, 2-year, and 4-year LDCTs), similarly to actual lung cancer screening practice. This setting allowed for testing the cumulative sensitivity of CAD in a longitudinal lung cancer screening with at least 3 time points. Among the SSNs overlooked by CAD at baseline, more than 50% of persistent SSN—namely, the prognostically relevant SSNs—were detected by CAD at subsequent LDCTs, without evidence of potential detrimental effect (eg, no interval cancer was seen in these SSNs). In particular, CAD showed a longitudinal 3-time-point sensitivity of 87.5%. Such findings need to be discussed within the context of patient management and potential overdiagnosis. The latter has been intensely debated for SSNs.[34,35] In our series, CAD as concurrent reader granted earlier detection of 3 of the 4 SSNs that grew into adenocarcinoma (Fig. 6); however, at the same time, it needs to be stated that they were resected only several years after their first documentation, without stage shift (data not reported). In the same context, it is conceivable that reliable detection of SSN plays a more important role as a general risk indicator rather than for the diagnosis of a slow growing primary lung tumor itself.[36,37]

Semiautomatic parameters (eg, diameter, volume, and mass) were not different in CAD-only and VD-only nodules in our study, indicating that size did not influence the CAD detection performance. Benzakoun et al[14] reported an inverse relation between whole diameter and CAD sensitivity; namely, the smaller the SSN, the higher the

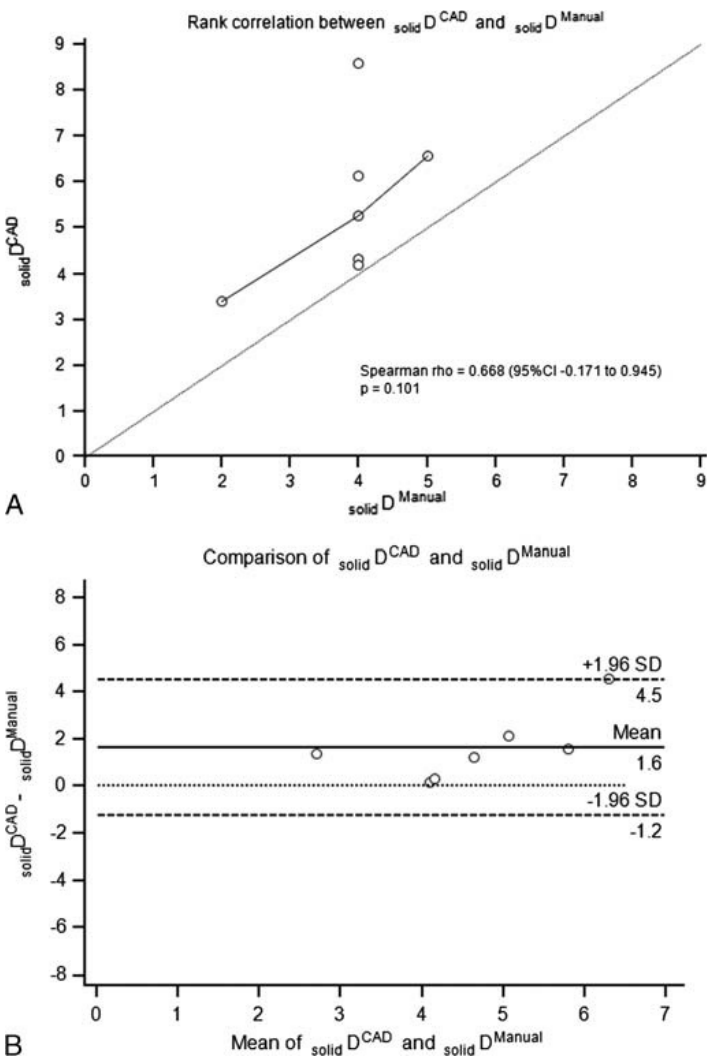

FIGURE 5. A–B, Scatter diagram of rank correlation (A) and Bland-Altman plot (B) for comparison between $_{solid}D^{CAD}$ and $_{solid}D^{Manual}$.

TABLE 3. Lung-RADS Category as Determined Either by Manual or Semi-automatic Measurement, and the Category Change

	All SSNs (n = 53)	NSN (n = 46)	PSN (n = 7)
Manual measurement			
Category 2	46 (87%)	46 (100%)	–
Category 3	7 (13%)	–	7 (100%)
Category 4A	–	–	–
Category 4B	–	–	–
Semiautomatic measurement			
Category 2	46 (87%)	46 (100%)	–
Category 3	4 (8%)	–	4 (57%)
Category 4A	2 (4%)	–	2 (29%)
Category 4B	1 (2%)	–	1 (14%)
Category change			
No change	50 (94%)	46 (100%)	4 (57.1%)
Increase by semiautomatic measurement	3 (6%)	–	3 (42.9%)
Reduced by semiautomatic measurement	–	–	–

The agreement was perfect for NSN. The category changed in 3 of 7 (43%) PSNs, notably with 3 PSNs classified Category 3 by manual measurement, which were upshifted to category 4A (n = 2) or 4B (n =1) by semiautomatic measurement.

Lung-RADS indicates Lung CT Screening Reporting and Data System; NSN, nonsolid nodule; PSN, part-solid nodule; SSN, subsolid nodule.

sensitivity of CAD. Yet, they highlighted that lowering the threshold of whole diameter caused a relevant increase of FP per scan with debatable clinical and prognostic implications.[14] A recent public challenge on automatic nodule detection showed that deep learning-based systems can outperform traditional CAD systems, so we expect that artificial intelligence will allow for further improvement of diagnostic performance.[32]

The second objective of our study was to compare semiautomatic and manual measurements of SSN diameters. We report that semiautomatic segmentation and manual measurement showed good correlation for the whole SSN diameter; only trivial systematic differences were seen. In that respect, the prototypical CAD used in our study seemed to be superior to previous reports that tested commercially available software on standard dose CT data.[14] Conversely, we observed a low correlation between semiautomatic segmentation and manual measurements of the solid component in PSN: CAD typically resulted in a larger diameter of the solid core (Fig. 7). With the limit of the small number of PSNs in our study, we observed a systematic shift toward a larger solid core diameter by semiautomatic measurement, potentially resulting in an up-shift classification according to Lung-RADS.[24] However, it has to be noted that the lack of histological proof does not allow for a final judgment of which method—manual or semiautomatic—would have correlated better with the invasive component of the tumor.[38,39] Cohen et al[39] reported that the histological measurement of invasive component was associated with the modulation of density threshold for the semiautomatic segmentation of solid core within PSN. Furthermore, future studies should include comparison of histology with more complex segmentation algorithms, such as machine learning, especially for the separation of solid core from the surrounding subsolid component and adjacent solid structures (eg, vessels).[40] Hence, CAD is suggested as concurrent or second reader for detection

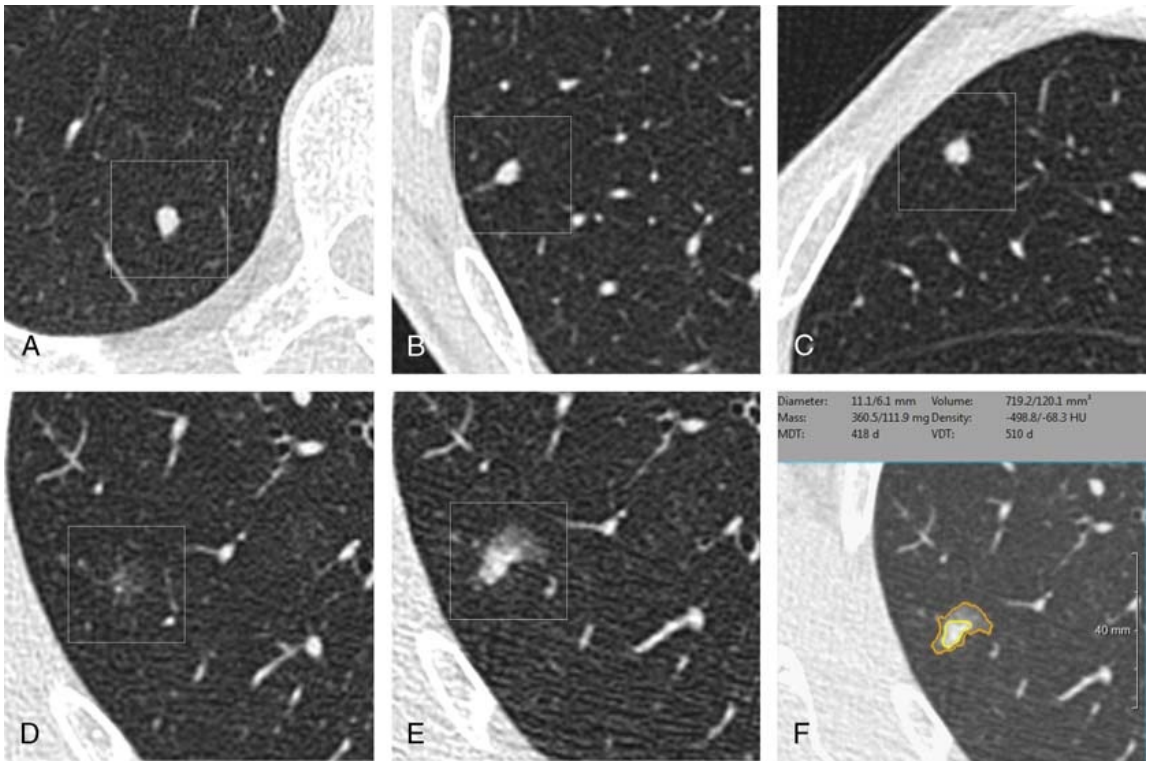

FIGURE 6. A–B, Axial noncontrast LDCT scan of a screenee with baseline NSN detected by CAD and overlooked by VD at baseline (60-year-old man). Baseline LDCT shows multiple solid nodules (A–C) and an NSN (D), the latter was detected by CAD and overlooked by VD. The presence of multiple solid nodules (>5 solid nodules) might have played a role in the overlooking of NSN by VD. The 4-year LDCT shows growth of NSN into a PSN (E) and its semiautomatic segmentation along with semiautomatic parameters (F—the orange contour outlines the whole segmented volume and the yellow contour outlines the solid core segmented volume); this nodule was resected with diagnosis of adenocarcinoma (stage Ia). MDT indicates mass doubling time; VDT, volume doubling time).

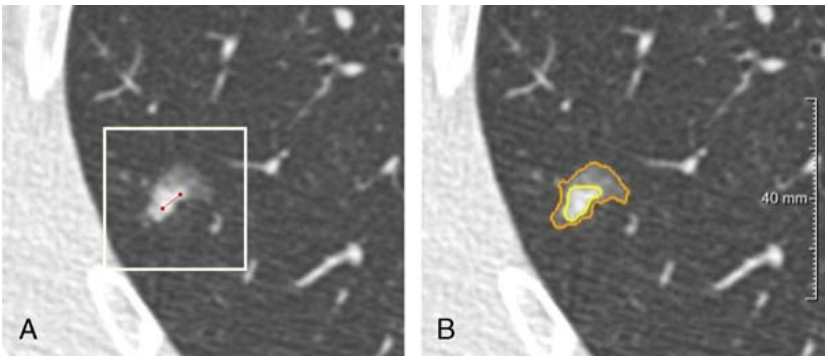

FIGURE 7. A–B, Axial noncontrast LDCT scan showing difference in the measurement of solid core either by manual or semiautomatic method. The manual measurement (A) of the diameter of solid core was 4 mm (red line; Lung-RADS category 3). The semiautomatic measurement (B) of the diameter of solid core was 6.1 mm (yellow outline; Lung-RADS category 4A).

of SSN in the routine reading, yet more histological confirmation is still granted before its application to nodule measurement.

Our study suffers from some limitations. First, the relatively small number of VD-only detected SSNs did not allow outlining a comprehensive signature of nodule features that are prone to CAD failure, notably for PSN. The small number of this group eventually limited the statistic power for such analysis. Second, all the LDCTs examined were performed by the same CT scanner, with consistent acquisition and reconstruction protocol. It should be noted that such ideal data feeding is not realistic for long-term lung cancer screening (eg, 55 through 80 years of age) because CT scanners and reconstruction algorithms are likely to be continuously updated. It is likely that performance of the tested CAD will vary with CT scanner features and reconstruction algorithms. Finally, the comparison of CAD with the cumulative visual performance of 7 radiologists might be seen as a limitation because of the inherent interobserver variability. On the other hand, this situation reflects the realistic screening scenario.

In conclusion, CAD provides complementary information to visual reading for the detection of SSN in volumetric LDCT for lung cancer screening. However, CAD marks require visual confirmation to correct for FP calls. The association of CAD and visual reading leads to optimal detection performance of SSN, especially of PSN. Further investigation is needed for segmentation of the solid component of PSN, ideally with histological reference.

ACKNOWLEDGMENTS

The authors thank Dr Alfonso Marchiano (chief of Radiology Department at IRCCS Istituto Nazionale Tumori, Milan), who granted prospective data collection of the MILD lung cancer screening trial.

REFERENCES

1. Henschke CI, Yankelevitz DF, Mirtcheva R, et al. CT screening for lung cancer: frequency and significance of part-solid and nonsolid nodules. *AJR Am J Roentgenol*. 2002;178:1053–1057.
2. Kim H, Park CM, Goo JM, et al. Quantitative computed tomography imaging biomarkers in the diagnosis and management of lung cancer. *Invest Radiol*. 2015;50:571–583.
3. Yip R, Henschke CI, Xu DM, et al. Lung cancers manifesting as part-solid nodules in the National Lung Screening Trial. *AJR Am J Roentgenol*. 2017;208:1011–1021.
4. Scholten ET, de Jong PA, de Hoop B, et al. Towards a close computed tomography monitoring approach for screen detected subsolid pulmonary nodules? *Eur Respir J*. 2015;45:765–773.
5. Hwang IP, Park CM, Park SJ, et al. Persistent pure ground-glass nodules larger than 5 mm: differentiation of invasive pulmonary adenocarcinomas from preinvasive lesions or minimally invasive adenocarcinomas using texture analysis. *Invest Radiol*. 2015;50:798–804.
6. Yankelevitz DF, Yip R, Smith JP, et al. CT screening for lung cancer: nonsolid nodules in baseline and annual repeat rounds. *Radiology*. 2015;277:555–564.
7. Gierada DS, Pinsky PF, Duan F, et al. Interval lung cancer after a negative CT screening examination: CT findings and outcomes in National Lung Screening Trial participants. *Eur Radiol*. 2017;27:3249–3256.
8. Zhao Y, de Bock GH, Vliegenthart R, et al. Performance of computer-aided detection of pulmonary nodules in low-dose CT: comparison with double reading by nodule volume. *Eur Radiol*. 2012;22:2076–2084.
9. Jeon KN, Goo JM, Lee CH, et al. Computer-aided nodule detection and volumetry to reduce variability between radiologists in the interpretation of lung nodules at low-dose screening computed tomography. *Invest Radiol*. 2012;47:457–461.
10. Rubin GD, Roos JE, Tall M, et al. Characterizing search, recognition, and decision in the detection of lung nodules on CT scans: elucidation with eye tracking. *Radiology*. 2015;274:276–286.
11. Rubin GD. Lung nodule and cancer detection in computed tomography screening. *J Thorac Imaging*. 2015;30:130–138.
12. Liang M, Tang W, Xu DM, et al. Low-dose CT screening for lung cancer: computer-aided detection of missed lung cancers. *Radiology*. 2016;281:279–288.
13. Scholten ET, Jacobs C, van Ginneken B, et al. Computer-aided segmentation and volumetry of artificial ground-glass nodules at chest CT. *AJR Am J Roentgenol*. 2013;201:295–300.
14. Benzakoun J, Bommart S, Coste J, et al. Computer-aided diagnosis (CAD) of subsolid nodules: evaluation of a commercial CAD system. *Eur J Radiol*. 2016;85:1728–1734.
15. Jacobs C, van Rikxoort EM, Murphy K, et al. Computer-aided detection of pulmonary nodules: a comparative study using the public LIDC/IDRI database. *Eur Radiol*. 2016;26:2139–2147.
16. Sverzellati N, Silva M, Calareso G, et al. Low-dose computed tomography for lung cancer screening: comparison of performance between annual and biennial screen. *Eur Radiol*. 2016;26:3821–3829.
17. Yoo RE, Goo JM, Hwang EJ, et al. Retrospective assessment of interobserver agreement and accuracy in classifications and measurements in subsolid nodules with solid components less than 8 mm: which window setting is better? *Eur Radiol*. 2017;27:1369–1376.
18. Jacobs C, van Rikxoort EM, Twellmann T, et al. Automatic detection of subsolid pulmonary nodules in thoracic computed tomography images. *Med Image Anal*. 2014;18:374–384.
19. Bankier AA, MacMahon H, Goo JM, et al. Recommendations for measuring pulmonary nodules at CT: a statement from the Fleischner Society. *Radiology*. 2017;285:584–600.
20. Jacobs C, van Rikxoort EM, Scholten ET, et al. Solid, part-solid, or non-solid?: classification of pulmonary nodules in low-dose chest computed tomography by a computer-aided diagnosis system. *Invest Radiol*. 2015;50:168–173.
21. Scholten ET, de Jong PA, Jacobs C, et al. Interscan variation of semi-automated volumetry of subsolid pulmonary nodules. *Eur Radiol*. 2015;25:1040–1047.
22. Travis WD, Asamura H, Bankier AA, et al. The IASLC Lung Cancer Staging Project: proposals for coding t categories for subsolid nodules and assessment of tumor size in part-solid tumors in the forthcoming eighth edition of the TNM Classification of Lung Cancer. *J Thorac Oncol*. 2016;11:1204–1023.
23. Weinstein S, Obuchowski NA, Lieber ML. Clinical evaluation of diagnostic tests. *AJR Am J Roentgenol*. 2005;184:14–19.
24. American College of Radiology website. 2014. Available at: https://www.acr.org/-/media/ACR/Files/RADS/Lung-RADS/LungRADS_AssessmentCategories.pdf. Accessed January 1, 2017.

25. Coblentz CL, Babcook CJ, Alton D, et al. Observer variation in detecting the radiologic features associated with bronchiolitis. *Invest Radiol*. 1991;26: 115–118.
26. Godoy MC, Kim TJ, White CS, et al. Benefit of computer-aided detection analysis for the detection of subsolid and solid lung nodules on thin- and thick-section CT. *AJR Am J Roentgenol*. 2013;200:74–83.
27. Yuan R, Vos PM, Cooperberg PL. Computer-aided detection in screening CT for pulmonary nodules. *AJR Am J Roentgenol*. 2006;186:1280–1287.
28. 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;237:657–661.
29. Young S, Lo P, Kim G, et al. The effect of radiation dose reduction on computer-aided detection (CAD) performance in a low-dose lung cancer screening population. *Med Phys*. 2017;44:1337–1346.
30. Gordic S, Morsbach F, Schmidt B, et al. Ultralow-dose chest computed tomography for pulmonary nodule detection: first performance evaluation of single energy scanning with spectral shaping. *Invest Radiol*. 2014;49:465–473.
31. Setio AA, Ciompi F, Litjens G, et al. Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. *IEEE Trans Med Imaging*. 2016;35:1160–1169.
32. Setio AAA, Traverso A, de Bel T, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge. *Med Image Anal*. 2017;42:1–13.
33. Yousaf-Khan U, van der Aalst C, de Jong PA, et al. Risk stratification based on screening history: the NELSON lung cancer screening study. *Thorax*. 2017;72: 819–824.
34. Patz EF Jr, Pinsky P, Gatsonis C, et al. Overdiagnosis in low-dose computed tomography screening for lung cancer. *JAMA Intern Med*. 2014;174:269–274.
35. Yip R, Wolf A, Tam K, et al. Outcomes of lung cancers manifesting as nonsolid nodules. *Lung Cancer*. 2016;97:35–42.
36. Detterbeck FC, Franklin WA, Nicholson AG, et al. The IASLC Lung Cancer Staging Project: background data and proposed criteria to distinguish separate primary lung cancers from metastatic foci in patients with two lung tumors in the forthcoming eighth edition of the TNM Classification for Lung Cancer. *J Thorac Oncol*. 2016;11:651–665.
37. Han S, Rivera GA, Cheng I, et al. PS01.77: risk-stratification for second primary lung cancer: topic: medical oncology. *J Thorac Oncol*. 2016;11:S319–S320.
38. Lee SM, Park CM, Song YS, et al. CT assessment-based direct surgical resection of part-solid nodules with solid component larger than 5 mm without preoperative biopsy: experience at a single tertiary hospital. *Eur Radiol*. 2017;27:5119–5126.
39. Cohen JG, Goo JM, Yoo RE, et al. Software performance in segmenting ground-glass and solid components of subsolid nodules in pulmonary adenocarcinomas. *Eur Radiol*. 2016;26:4465–4474.
40. Charbonnier JP, Chung K, Scholten ET, et al. Automatic segmentation of the solid core and enclosed vessels in subsolid pulmonary nodules. *Sci Rep*. 2018;8:646.