



Consistency of radiologists in identifying pulmonary nodules based on low-dose computed tomography

Shuai Ming, Wei Yang, Si-Jia Cui, Shuai Huang, Xiang-Yang Gong

Department of Radiology, Zhejiang Provincial People's Hospital, Affiliated People's Hospital of Hangzhou Medical College, Hangzhou 310014, China

Contributions: (I) Conception and design: XY Gong, S Ming; (II) Administrative support: None; (III) Provision of study materials or patients: S Ming; (IV) Collection and assembly of data: W Yang, SJ Cui, S Huang; (V) Data analysis and interpretation: S Ming; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Xiang-Yang Gong. Department of Radiology, Zhejiang Provincial People's Hospital, Affiliated People's Hospital of Hangzhou Medical College, Institution of Teleradiology and Artificial Intelligence of Hangzhou Medical College, Hangzhou 310014, China. Email: cjr.gxy@hotmail.com.

Background: To study the consistency of radiologists in identifying pulmonary nodules based on low-dose computed tomography (LDCT), and to analyze factors that affect the consistency.

Methods: A total of 750 LDCT cases were collected randomly from three medical centers. Three experienced chest radiologists independently evaluated and detected the pulmonary nodules on 625 cases of LDCT images. The detected nodules were classified into 3 groups: group I (detected by all radiologists); group II (detected by two radiologists); group III (detected by only one radiologist). The consistency with respect to the image features of individual nodules was assessed.

Results: A total of 1,206 nodules were identified by the three radiologists. There were 234 (19.4%) nodules in group I, 377 (31.3%) nodules in group II, and 595 (49.3%) nodules in group III. Logistic regression showed that the size, density, and location of the nodules correlated with the detection of nodules. Nodules sized great than or equal to 4 mm were more consistently identified than nodules sized less than 4 mm. Solid and calcified nodules were more consistently identified than sub-solid nodules. Nodules located in the outer zone were more consistently identified than hilar nodules.

Conclusions: There was considerable inter-reader variability with respect to identification of pulmonary nodules in LDCT. Larger nodules, solid or calcified nodules, and nodules located in the outer zone were more consistently identified.

Keywords: Pulmonary nodules; low-dose computed tomography (LDCT); radiologist; inter-reader variability

Submitted Mar 16, 2019. Accepted for publication Jul 02, 2019.

doi: 10.21037/jtd.2019.07.52

View this article at: <http://dx.doi.org/10.21037/jtd.2019.07.52>

Introduction

The use of low-dose computed tomography (LDCT) for screening of pulmonary nodules in at-risk population was shown to reduce lung cancer mortality by 20% as compared to screening with chest radiography (1). An estimated 4.8 million Americans are believed to have undergone at least one chest LDCT scan and 1.57 million were found to have nodules; of these, 63,000 were diagnosed with lung

cancer within a period of 2 years (2). Lung cancer accounted for 17.1% of all newly diagnosed cancer cases 21.7% of all cancer-related mortality in China (3). Therefore, early screening and treatment became a top priority for lung cancer (4). The experience of radiologists plays a major role in the detection and correct categorization of the pulmonary nodules on screening LDCT (5).

Although, there have many guidelines for reporting of pulmonary nodules, radiologists still have a huge difference

in judging specific pulmonary nodules. Objectively, there are some nodules that are easily overlooked, especially for hilar nodules and ground-glass opacity nodules (6,7). In a study with LDCT, more than one third of nodules were missed at the baseline scan, and were detected at the follow-up scan 1 year later (8). A consensus double reading did increase the pulmonary nodules detectability of 19%, compared with a single reading (9).

There were a few studies dealing with the inter-reader variability of pulmonary nodules on CT screening (10-12). Inter-observer variability performance for the detection of pulmonary nodules has been found to be relatively high with both standard-dose and LDCT (13). However, the article on the analysis of causes of inconsistencies between radiologists has not yet been published. Our aim was to evaluate the consistency of radiologists in judging pulmonary nodules and its correlation with image features of the pulmonary nodules.

Methods

Imaging data

This retrospective study was approved by the Ethics Committee of Zhejiang Provincial People's Hospital. The requirement for written informed consent of patients was waived off because of retrospective use of imaging data.

A total of 730 chest LDCT cases were randomly collected from three different medical centers (Zhejiang Provincial People's Hospital, Zhejiang Hospital and Tongde Hospital) from the PACS system between June to July 2017. All patient sensitive information was eliminated from the data. Two radiology residents screened the 730 LDCT cases and eliminated 75 cases based on the following exclusion criteria: incomplete images; presence of motion or metal artifacts; history of chest surgery; diffuse pulmonary disease; presence of more than 10 pulmonary nodules. Then, 30 cases were randomly selected from the remaining 655 cases and reviewed together by three senior radiologists (each having more than 20 years' experience) to familiarize with the operating system and reach a common sense of criteria.

Equipment and examination parameters

Chest CT images were acquired with a low-dose CT protocol on 64-slice multi-detector CT scanners (Somatom Definition AS, AS⁺ & Flash, Siemens, Germany; Optima

680, Discovery 750 HD, GE Healthcare, USA). The following CT protocol settings were used: 120 kV; 10-40 mA; gantry rotation speed: 0.5 seconds; helical scan mode (cranial-caudal direction) with a pitch of 1.2. The cumulative radiation dose was <1 mGy for each patient. Breath-holding lasted for 15 seconds to prevent image motion artifacts during the scan. Lung LDCT images were then reconstructed using a kernel of B75f at a slice thickness of 2 mm. All image readings were performed at a lung window width of 1,500 HU and a window level of -400 HU.

A dedicated system for lung nodule assessment was used. Professional reading monitors (Giant Shark, Nanjing Giant Shark Medical) were used to display LDCT images.

Evaluation of pulmonary nodules

Three experienced radiologists evaluated the 625 lung LDCT cases independently with no restrictions of time. Pulmonary nodules detected by each radiologist were recorded automatically. All detected nodules were categorized into three groups: nodules detected by all three radiologists (group I); nodules detected by two of the three radiologists (group II); and nodules detected by only one radiologist (group III). The flowchart was shown in *Figure 1*.

Two chest radiologists reviewed all nodules one by one, and recorded imaging features of each nodule presented by former three radiologists. Another expert made a final judgment in case of any controversy.

The following image features of the nodules were analyzed: size, density and location. The nodules were divided into two groups based on the diameter: less than 4 mm and greater than or equal to 4 mm (14). The location of the nodules was classified as follows: a peripheral nodule was defined as a nodule located within 2 cm of the pleura. A hilar nodule was located within 2 cm of the hilus. Central nodule was located between the peripheral and hilar zones (15). The nodules were classified as solid, subsolid, or calcified according to the nodule attenuation (16) (*Figures 2-4*).

Statistical analysis

Normally distributed continuous variables are expressed as mean \pm standard deviation or as median and interquartile range (IQR), as appropriate. Categorical variables are expressed as percentage. Chi-square test was used to

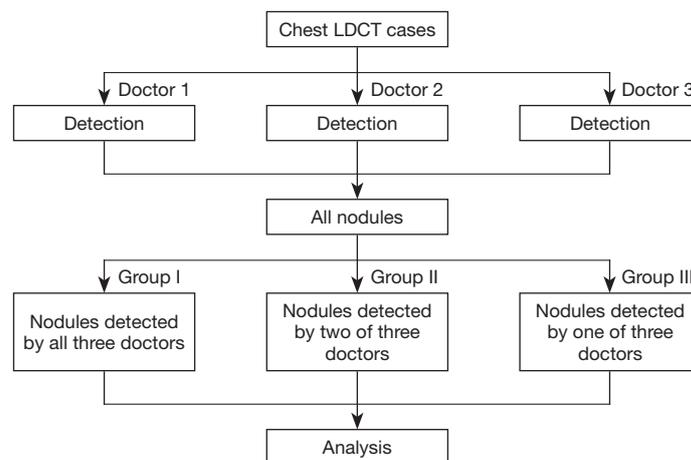


Figure 1 Flowchart for detecting pulmonary nodules. LDCT, low-dose computed tomography.

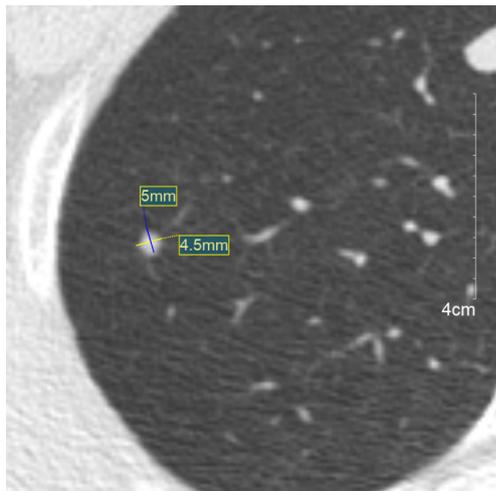


Figure 2 Nodules sized >4 mm.

assess between-group differences with respect to nodule characteristics. Multiple logistic regression analysis was used to analyze the correlation between radiologist's consistency and nodule characteristics. All analyses were performed using SPSS (version 18.0; Chicago, III). Statistical significance level was set at $P < 0.05$.

Results

Descriptive statistics

A total of 625 chest LDCT cases were included in this study. The mean age of patients was 50.67 ± 13.52 years [range, 20–84 years; median age: 63 years (interquartile

range, 43–78 years)].

A total of 1,206 nodules were detected with an average of 1.9 (1,206/625) nodules per case. There were 234 (19.4%) nodules in group I, 377 (31.3%) nodules in group II, and 595 (49.3%) nodules in group III (Figure 5). Chi-square analysis revealed a significant correlation between groups and nodular parameters ($P < 0.001$) (Table 1).

Nodule size

Among the 1,206 nodules, 1,070 (88.7%) were less than 4 mm, and only 136 (11.3%) were greater than or equal to 4 mm. There was a significant difference with respect to nodule size between groups I and II ($P < 0.001$) and between groups I and III ($P < 0.001$). No significant difference was observed between groups II and III in this respect ($P = 0.765$).

Nodule location

There were 37 (3.1%) nodules at the hilar zones, 188 (15.6%) at the central zone, 981 (81.3%) peripheral nodules. There was significant difference between group II and III ($P < 0.001$) and between groups I and III ($P < 0.001$) in terms of location. No significant difference was observed between groups I and II in this respect ($P = 0.240$).

Nodule density

A total of 266 (22.1%), 658 (54.6%) and 282 (23.4%) nodules were classified as subsolid, solid and calcified, respectively. There were significant differences with respect to nodule density type between groups I and II, groups I and III, groups II and III ($P < 0.001$).

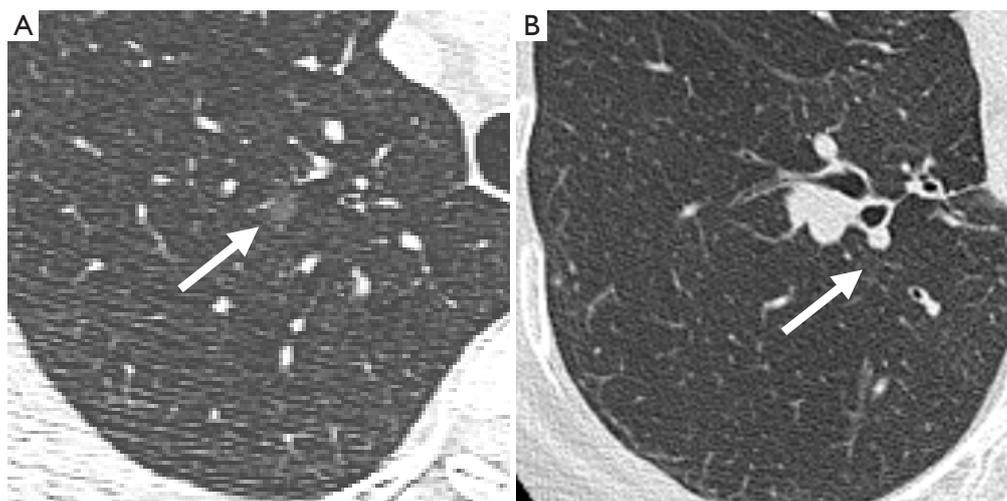


Figure 3 Nodules of inconsistently detected. (A) A 49-year-old female, a subsolid nodule (arrow); (B) a 73-year-old female, a hilar nodule (arrow).

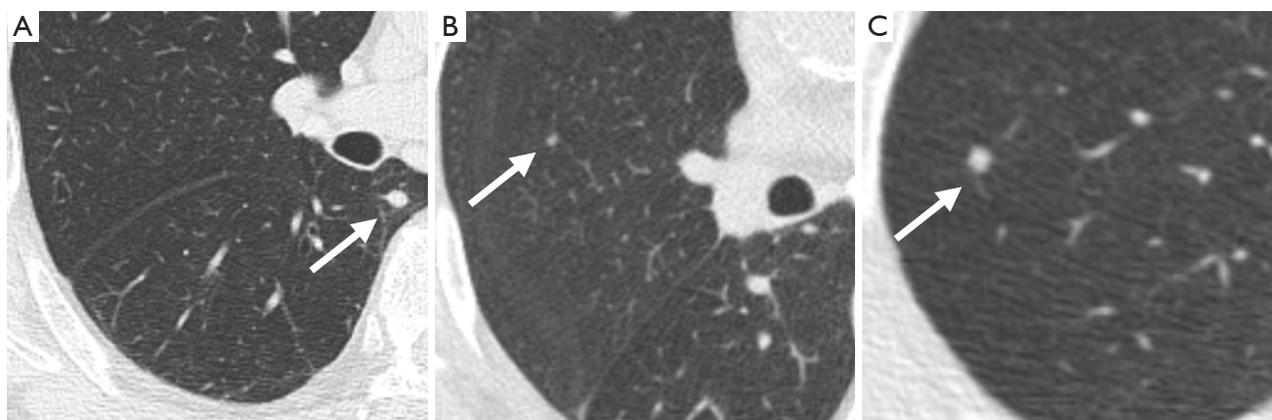


Figure 4 Nodules of consistently detected. (A) A 63-year-old male, a hilar nodule (arrow); (B) an 82-year-old male, a central nodule (arrow); (C) a 54-year-old male, peripheral nodule (arrow).

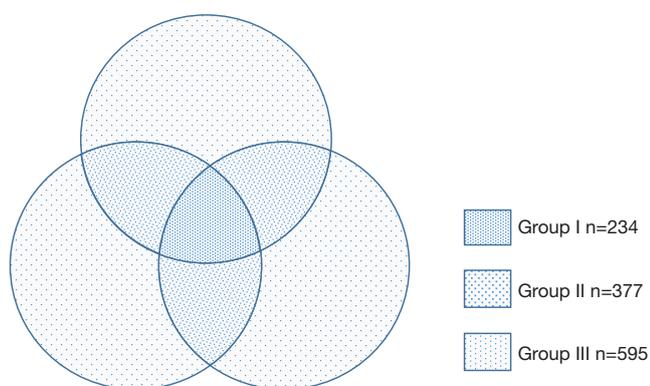


Figure 5 All nodules groups depend on the consistence of doctors.

Multiple logistic regression analysis

We selected group I, which was detected by all radiologists, as the positive group and group III was the negative group. The two groups were analyzed by multivariate logistic regression analysis. Logistic regression analysis of nodule size yielded an odds ratio of 0.053 [95% confidence interval (CI): 0.031–0.093; $P < 0.001$]. There was a significant difference between categorization of central and peripheral nodules ($P < 0.001$). There was a significant difference with respect to categorization of solid and calcified nodules ($P < 0.001$) (Table 2).

A sub-analysis was performed for the nodules greater

Table 1 The correlation between consistency and imaging features of pulmonary nodules

Features	Group I	Group II	Group III	Total	χ^2	P
Size, mm					194.737	<0.001
<4	147	357	566	1,070		
≥4	87	20	29	136		
Location					57.336	<0.001
Hilar	1	6	30	37		
Central	19	40	129	188		
Peripheral	214	331	436	981		
Density					139.039	<0.001
Subsolid	40	79	147	266		
Solid	118	105	347	658		
Calcification	76	193	101	282		
Total	234	377	595	1,206		

Group I represents agreements of three, group II represents agreements of two, group III represents agreements of only one.

Table 2 A multivariate logistic regression between group I and group III

Factors	B	SE	Wals	df	Sig	Exp(B)	95% CI for EXP (B)
Size	-2.930	0.283	106.914	1	0.000	0.053	0.031–0.093
Hilar			36.937	2	0.000		
Central	-3.918	1.078	13.218	1	0.000	0.020	0.002–0.164
Peripheral	-1.582	0.308	26.425	1	0.000	0.206	0.112–0.376
Subsolid			22.598	2	0.000		
Solid	-1.085	0.269	16.283	1	0.000	0.338	0.200–0.572
Calcification	-0.919	0.214	18.414	1	0.000	0.399	0.262–0.607
Constants	2.546	0.326	61.123	1	0.000	12.760	

The subsolid nodule and hilar nodule as the reference categories.

than or equal to 4 mm. Group I was selected as positive group and group III as negative group. The two groups were analyzed by multivariate logistic regression analysis. There was a significant difference between categorization of central and peripheral nodules ($P<0.001$). There was no significant difference between categorization of solid nodules and calcified nodules ($P=0.292$) (Table 3).

Discussion

Pulmonary nodules are commonly detected on chest CT examination, with an incidence varies from 13% to

58% (7-10). The sensitivity of radiologists in detecting pulmonary nodules ranges from 64% to 82% based on LDCT (9,10). Due to the lack of gold standard, pulmonary nodules are typically judged by two or more doctors as a reference standard. In addition to the actual incidence of pulmonary nodules, doctors' different detection and reporting of pulmonary nodules may be one of the possible reasons for the great variation. Samim *et al.* (17) reported the agreement between two radiologists with respect to identification of pulmonary nodules was 0.78 kappa value in patient-based analysis and 0.40 in lobe-based analysis. In the present study, only 19.4% of 1,206 nodules were

Table 3 A sub-analysis multivariate logistic regression between group I and group III

Factor	B	SE	Wals	df	Sig	Exp (B)	95% CI for EXP (B)
Hilar			24.131	2	0.000		
Central	-3.939	1.134	12.057	1	0.001	0.019	0.002-0.180
Peripheral	-2.331	0.577	16.293	1	0.000	0.097	0.031-0.302
Subsolid			1.808	2	0.405		
Solid	-1.016	0.924	1.208	1	0.272	0.362	0.059-2.216
Calcification	-1.015	0.772	1.731	1	0.188	0.362	0.080-1.644
Constants	2.836	0.718	15.598	1	0.000	17.043	

The subsolid nodule and hilar nodule as the reference categories.

recognized by all three raters, although all three radiologists were experienced and underwent a common training before the study.

A previous study by Wormanns *et al.* (13) showed that the consistency of pulmonary nodules assessed by three radiologists was 47%. Our result presented a lower consistency of 19%. There were two differences between the two studies. First, we used clinical data, which include positive and negative cases, and Wormanns' entire case was clearly positive in advance to readers; second, we use the number of more cases, radiologists need to pay more workload in the assessment.

This study explored the relationship between the consistency of pulmonary nodules detection and its image features. The results showed that the size, location and density of pulmonary nodules significantly affected the consistency of radiologists large diameter nodules, solid, calcified nodules and peripherally distributed pulmonary nodules have higher consistency, because these nodules are more easily detected by radiologists than small nodules, sub-solid nodules and inner distributed nodules, which is consistent with previous research and our work experience.

On sub-analysis for larger than or equal to 4 mm nodules, our results showed no significant difference in consistency among radiologists about the density of nodules. It means the consistency is less affected by the density for the larger nodules. However, the location was still the factor for the consistency.

We use 4 mm diameter as a threshold according the guidelines (18-24). We referred to the 2016 National Comprehensive Cancer Network guidelines (NCCN), 2017 Fleischer Society guidelines, the 2013 American College of Chest Physicians guidelines (ACCP) and the 2016 Clinical practice consensus guidelines for Asia. Four millimeters was

the minimum threshold for nodules do not require routine follow-up.

In the present study, most nodules smaller than 4 mm were detected by only one radiologist. We speculate that radiologists tend to ignore small nodules which appear clinically insignificant. This selectivity exhibits greater randomness, which may be the main reason for the difference of detecting results. However, there were still some pulmonary nodules larger than 4mm missing by radiologist. Most of these nodules were in the hilar location or subsolid nodules. This shows that even if experienced radiologists, the possibility of missed detection still exists. Check by a second or third radiologist may help improve the accuracy of the diagnostic report. The artificial intelligence, which has developed rapidly in recent years, shows high sensitivity and accuracy to pulmonary nodules (25-27). AI's pre-processing of images may become an assistant of radiologists in pulmonary detection.

There were several limitations in this study. First, none of the detected nodules were confirmed by surgery and pathology. Therefore, the total number of true nodules remains unknown. Second, the three radiologists belonged to different medical centers; although consistent training was carried out, there may still be differences in diagnostic habits. Third, the LDCT images in our study population were obtained with 2 mm section thickness, not the 1.25 mm from the guideline.

Conclusions

There was considerable inter-reader variability with respect to characterization of pulmonary nodules in LDCT. Larger nodules, solid or calcified nodules, and nodules in the outer zone were more likely to be consistently evaluated.

Acknowledgments

None.

Footnote

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

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This retrospective study was approved by the Ethics Committee of Zhejiang Provincial People's Hospital. The requirement for written informed consent of patients was waived off because of retrospective use of imaging data.

References

1. Aberle DR, Adams AM, Berg CD, et al. Reduced Lung Cancer Mortality with Low-dose Computed Tomographic Screening. *N Engl J Med* 2011;365:395-9.
2. Gould MK, Tang T, Liu L, et al. Recent Trends in the Identification of Incidental Pulmonary Nodules. *Am J Respir Crit Care Med* 2015;192:1208-14.
3. Zhi XY, Zou XN, Hu M, et al. Increased Lung Cancer Mortality Rates in the Chinese Population from 1973-1975 to 2004-2005: An Adverse Health Effect from Exposure to Smoking. *Cancer* 2015;121:3107-12.
4. Gupta A, Saar T, Martens O, et al. Automatic Detection of Multisize Pulmonary Nodules in CT images: Large scale Validation of the False-positive Reduction Step. *Med Phys* 2018;45:1135-49.
5. Dou Q, Chen H, Yu L, et al. Multilevel Contextual 3-D CNNs for False Positive Reduction in Pulmonary Nodule Detection. *IEEE Trans Biomed Eng* 2017;64:1558-67.
6. Wang Y, van Klaveren RJ, de Bock GH, et al. No Benefit for Consensus Double Reading at Baseline Screening for Lung Cancer with the Use of Semiautomated Volumetry Software. *Radiology* 2012;262:320-6.
7. Lee KH, Goo JM, Park SJ, et al. 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* 2014;9:74-82.
8. Swensen SJ, Jett JR, Sloan JA, et al. Screening for Lung Cancer with Low-dose Spiral Computed Tomography. *AM J Respir Crit Care Med* 2002;165:508-13.
9. Martini K, Barth BK, Nguyen-Kim TD, et al. Evaluation of Pulmonary Nodules and Infection on Chest CT with Radiation dose Equivalent to Chest Radiography: Prospective Intra-individual Comparison Study to Standard Dose CT. *Eur J Radiol* 2016;85:360-5.
10. Schreuder A, van Ginneken B, Scholten ET, et al. Classification of CT Pulmonary Opacities as Perifissural Nodules: Reader Variability. *Radiology* 2018;288:867-75.
11. van Riel SJ, Sánchez CI, Bankier AA. Observer Variability for Classification of Pulmonary Nodules on low-dose CT Images and its Effect on Nodule Management. *Radiology* 2015;277:863-71.
12. Wiener RS, Gould MK, Slatore CG, et al. Resource use and guideline concordance in evaluation of pulmonary for cancer: too much and too little care. *JAMA Intern Med* 2014;174:871-80.
13. Wormanns D, Ludwig K, Beyer F, et al. Detection of Pulmonary at Multislice-detector CT: Effectiveness of Double Reading to Improve Sensitivity at Standard-dose and Low-dose Chest CT. *Eur Radiol* 2005;15:14-22.
14. Shen H. Low-dose CT for Lung Cancer Screening: Opportunities and Challenges. *Front Med* 2018;12:116-21.
15. Yuan R, Vos PM, Cooperberg PL, et al. Computer-Aided Detection in Screening CT for Pulmonary Nodules. *AJR Am J Roentgenol* 2006;186:1280-7.
16. Choi WS, Park CM, Song YS, et al. Transient Subsolid Nodules in Patients with Extrapulmonary Malignancies: their Frequency and Differential Features. *Acta Radiol* 2015;56:428-37.
17. Samim A, Littooj AS, van den Heuvel-Eibrink MM, et al. Frequency and Characteristics of Pulmonary Nodules in Children at Computed Tomography. *Pediatr Radiol* 2017;47:1751-8.
18. Lacson R, Prevedello LM, Andriole KP, et al. Factors Associated with Radiologist Adherence to Fleischner Society Guidelines for management of Pulmonary Nodules. *J Am Coll Radiol* 2012;9:468-73.
19. Wang YX, Gong JS, Suzuki K, et al. Evidence Based Imaging Strategies for Solitary Pulmonary Nodule. *J Thorac Dis* 2014;6:872-87.
20. Wood DE. National Comprehensive Cancer Network (NCCN) Clinical Practice Guidelines for Lung Cancer Screening. *Thorac Surg Clin* 2015;25:185-97.
21. Nair A, Devaraj A, Callister MEJ, et al. The Fleischner Society 2017 and British Thoracic Society 2015 Guidelines for Managing Pulmonary Nodules: Keep Calm and Carry on. *Thorax* 2018. doi 10.1136/thoraxjnl-2018-212393.

22. MacMahon H, Naidich DP, Goo JM, et al. Guidelines for Management of Incidental Pulmonary Nodules Detected on CT Images: From the Fleischner Society 2017. *Radiology* 2017;284:228-43.
23. Bai C, Choi CM, Chu CM, et al. Evaluation of Pulmonary Nodules: Clinical Practice Consensus Guidelines for Asia. *Chest* 2016;150:877-93.
24. Gould MK, Donington J, Lynch WR, et al. Evaluation of Individuals with Pulmonary Nodules: When is it Lung Cancer? *Diagnosis and Management of Lung Cancer*, 3rd ed. American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest* 2013;143:93-120.
25. Setio AA, Jacobs C, Gelderblom J, et al. Automatic Detection of Large Pulmonary Solid Nodules in Thoracic CT Images. *Med Phys* 2015;42:5642-53.
26. Valente IR, Cortez PC, Neto EC, et al. Automatic 3D Pulmonary Nodule Detection in CT Images: A survey. *Comput Methods Programs Biomed* 2016;124:91-107.
27. Landreneau RJ, Schuchert MJ. Is Segmentectomy the Future? *J Thorac Dis* 2019;11:308-18.

Cite this article as: Ming S, Yang W, Cui SJ, Huang S, Gong XY. Consistency of radiologists in identifying pulmonary nodules based on low-dose computed tomography. *J Thorac Dis* 2019;11(7):2973-2980. doi: 10.21037/jtd.2019.07.52